MASTER EN SCIENCES INFORMATIQUES

Reconnaissance faciale des émotions
Un regard multidisciplinaire

Caluwaerts, Michel

Award date:
2018

Awarding institution:
Universite de Namur

Université de Namur
Faculté d'informatique
Année académique 2017–2018

Reconnaissance faciale des émotions
Un regard multidisciplinaire

Michel Caluwaerts

Maître de stage :

Promoteur : _____ (Signature pour approbation du dépôt - REE art. 40)
Claire Lobet-Maris

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Informatiques.

Préface

Nous tenons à remercier chaleureusement Madame Lobet-Maris dont le soutien a été constant, tout au long de l'écriture de ce travail. Face à un sujet qu'il nous a fallu explorer tous les deux, ses conseils, sa pédagogie et sa disponibilité, même pour relire en dernière minute, nous ont été précieux. Dans les moments de grand découragement, il suffit parfois de quelques mots pour repartir à l'attaque et ceux là ont été décisifs pour nous.

Je ne trouve pas assez de mots non plus pour remercier mes proches Valérie-Anne en particulier, toi qui as fait tourner la maison pratiquement seule pendant 3 mois, et vous deux Simon et Lionel qui m'ont toujours témoigné votre affection malgré le peu de disponibilité que j'avais pour vous.

Enfin nous remercions nos parents pour avoir été là, silencieusement mais chaque fois prêts à aider, surtout au plus près des échéances.

Rien de tout ceci n'aurait vu le jour sans vous.

Résumé

La reconnaissance faciale des émotions est un domaine de recherche à l'intersection des champs de la recherche en psychologie, en sciences comportementales et en informatique. Nos indices émotionnels sont convoités par les GAFA[1] pour enrichir leur modèle publicitaire et nous proposer des contenus plus contextuels. L'industrie de la surveillance compte sur l'extension de la vidéo-surveillance à la reconnaissance des émotions pour adjoindre à nos identités numériques des indices sur nos comportements ou nos tendances à l'action. Dans ce mémoire, nous proposons un état de l'art de la reconnaissance faciale de l'émotion. Nous présentons et comparons d'abord les principaux modèles théoriques de l'émotion. Pour illustrer comment le modèle des émotions de base issu des travaux de Paul Ekman s'est imposé dans la construction des systèmes informatique, nous détaillons les principes de construction des systèmes de reconnaissance faciale de l'expression et des bases de données sur lesquelles ils sont entraînés. Comme alternative au modèle des émotions de base, nous présentons le modèle de l'*appraisal* et nous les comparons. Nous terminons par une critique de la validité du concept de reconnaissance des émotions de base, portant tant sur ses fondements théoriques que sur la construction de sa base expérimentale.

Mots-clé : émotion, surveillance, reconnaissance faciale des emotions, théorie des émotions, Ekman, appraisal.

Abstract

Facial emotion recognition, although yet in a developping phase, has been a growing trend for the last decades, challenging a multidisciplinary body of research. The emotional cues on our faces are of the highest interest for the GAFA-like industry giants, seeking to enrich their contents with user-contextual data. Actors of surveillance and state security matters challenge the techonology, with the expectation that an « augmented »facial recognition, able to discern our emotional affects would allow them to conduct better behaviour analysis and prediction. In this thesis, we first review the traditional models in the theories of emotions. We then illustrate the most dominant of them, originating from psychologist Paul Ekman's works, and depict how it currently governs most of the conceptual construction principles of existing systems. We also review the contruction methodologies of training datasets, and codification schemes. As a conclusion, we assess the scientific validity of the concept of emotion recognition by questionning theoretic foundations and ground thruth construction.

Keywords :Facial expression recognition, emotions, facial recognition, theory of emotions, Ekman, dataset, facial muscle detection.

1. Google, Apple, Facebook, Amazon.

Table des matières

Table des figures

Liste des tableaux

Acronymes

ATF Appraisal Tendency Framework.
AU Action Unit.

CPM Component Process Model.

EMFACS Emotional facial action coding system.

FACS Facial Action Coding System.
FAP Facial Animation Parameters.
FAPU FAP Unit.
FAU Facial Action Unit.
FERA Facial Expression Recognition Challenge.
FP Feature Point.

GAFA Google, Apple, Facebook, Amazon.

HCI Human-Computer Interaction.

IHM Interaction Homme-Machine.
ISFER Integrated System for Facial Expression Recognition.

JAFFE Japanese Female Facial Expression.

MPEG Moving Pictures Experts Group.
MU Motion Unit.

NBN Neural Bayesian Network.
NIST National Institute of Standards and Technologies.
NN Neural Network.

SEC Stimulus Evaluation Check.
SNHC Synthetic/Natural Hybrid Coding.
SVM Support Vector Machines.

Introduction

Autrefois considérées comme un fardeau mental obscurcissant le jugement rationnel, l'émotion est aujourd'hui approchée comme un élément essentiel de nos interactions sociales. Nos émotions nous servent à colorer nos relations, elles appuient nos comportements, annoncent nos intentions, et on sait grâce aux neuro-sciences qu'elle sont un facteur déterminant de notre capacité à agir.

Si nos émotions peuvent se manifester à travers les différents canaux que sont le langage corporel, le timbre de voix ou les signaux physiologiques, c'est le visage qui en est le vecteur le plus efficace [Mehrabian, 2008].

Il n'est donc pas surprenant qu'un large champ de recherche se soit développé, d'abord dans les sciences affective et comportementales, ensuite en sciences informatiques, pour faire du visage une interface entre notre expérience affective interne et le monde extérieur.

Depuis les travaux de Charles Darwin [Darwin, 1872] qui, dans la lignée de sa Théorie de l'Évolution, identifie six états émotionnels fondamentaux, différentes fonctions ont été données à nos émotions par les sciences humaines. Phénomène adaptatif, outil de régulation sociale de nos comportements, source de tendance à l'action, plus d'un siècle de recherche n'ont pas encore abouti à arrêter un consensus sur ce que *sont*, et ce que *font* réellement nos émotions.

Avec le développement des technologies de l'information, le déchiffrement de nos expressions faciales a pris une autre dimension. Exploiter les techniques de reconnaissance faciale pour en extraire les émotions ouvre des perspectives nouvelles dans les domaines économique, sociétal et industriel. Les émotions sont devenues une matière première dont la maîtrise est très convoitée.

Nous illustrons le contexte dans lequel naît cet engouement par 3 éléments : l'économie de l'attention, le développement de nouvelles techniques de surveillance et la concurrence exacerbée entre les Google, Apple, Facebook, Amazon (GAFA) pour la création de valeur autour de contenus enrichis.

Reprenant un concept de la fin du XX$^{\text{ème}}$ siècle formulé par Goldhaber [Goldhaber, 1997], l'économie de l'attention [Kessous et al., 2010] propose une modélisation des échanges économiques basée sur une ressource centrale qui est l'attention.

Dans un monde où la quantité d'information augmente de manière exponentielle, où nous sommes exposés à un nombre croissant de média au travers d'outils informatiques proposant des contenus toujours plus riches, c'est l'information qui chasse l'attention de son public. Dans

le modèle de l'attention, les rapports économiques entre producteurs et consommateurs d'information sont repensés. Notre attention est devenue une ressource rare que le modèle économique publicitaire se doit d'exploiter au mieux, dans un contexte de concurrence féroce entre un nombre décroissant de grand acteurs et un nombre croissant d'internautes.

Remplaçant le paradigme de l'équilibre par l'offre et la demande de biens en économie classique, un modèle de l'attention fait l'hypothèse d'émetteurs chassant des récepteurs d'informations, ces derniers se tournant naturellement vers les émetteurs les plus puissants.

Capturer l'attention des internautes dans un monde où l'information est pléthorique suppose de pouvoir le « saisir »au plus près de ce qu'il est, et de ce qu'il ressent.

Dans ce cadre, les émotions sont devenues un puissant vecteur d'approche de ces internautes afin de les profiler et leur adresser du contenu média ou publicitaire correspondant à leur profil émotionnel.

Le déploiement à grande échelle déjà, depuis quelques décennies, d'une infrastructure de surveillance vidéo ouvre un autre domaine d'application à l'analyse de l'expression faciale.

La biométrie en général répond à un besoin des services de sécurité de plus en plus marqué de recourir à la technologie, qui seule, par son efficacité, permet de « toujours précéder la progression de la menace »[Ceyhan, 2006] dans la lutte contre une menace toujours plus diffuse.

Dans un monde plus global et plus ouvert, où les individus sont de plus en plus mobiles, les technologies de reconnaissance faciale des émotions sont de plus en plus utilisées. S'il devient possible d'analyser en temps réel l'expression portées par un visage, il devient alors possible d'en retirer des données biométriques plus riches, une reconnaissance faciale « augmentée ».

L'hypothèse que l'on peut par la suite en inférer à tout moment un état affectif, et de là, en déduire une tendance à l'action, répond justement à ce besoin d'anticiper la menace.

En conclusion, ces systèmes font l'hypothèse que l'analyse des émotions permet de prédire ou détecter la dangerosité potentielle d'un individu.

L'économie de l'attention et la surveillance expliquent que dans la concurrence acharnée que se livrent les géants de l'industrie numérique, l'exploitation des émotions soit devenu un enjeu stratégique.

En témoignent les récentes acquisitions effectuées par les grands acteurs, visant à acquérir la technologie : Emotient racheté par Apple, FacioMetrics par Facebook, Orbeus par Amazon, etc. Toutes ces startups, ont en commun une expertise dans le travail de reconnaissance de l'expression faciale.

En se dotant de la technologie de reconnaissance des émotions, ces firmes entendent se démarquer en proposant des contenus plus interactifs, mais aussi et surtout, exploiter en temps réel l'état émotionnel de l'utilisateur afin de proposer à tout moment un contenu média ou publicitaire adapté au contexte émotionnel de la cible.

Pour une entreprise comme Google dont le moteur économique est sa puissante régie publicitaire, il est stratégique de pouvoir proposer, en plus d'un espace publicitaire, un ciblage et une rétro-analyse qualitative des réactions émotionnelles qu'il suscite sur les récepteurs.

L'objectif de ce mémoire se veut exploratoire. Il s'agit de comprendre sur quelles bases théoriques et techniques reposent les systèmes de reconnaissance faciale de l'émotion.

Que sont les émotions, de quoi témoignent-elles, comment peut-on les détecter, et comment il est possible de les codifier. Tel est le cheminement que nous avons suivi.

Le travail que nous présentons se veut être un état de l'art non exhaustif de la technologie de classification de l'imagerie faciale, mais ayant plutôt pour but de comprendre la construction théorique des émotions, du lien entre expression faciale et émotion, et des choix méthodologiques sur lesquels sont fondés les systèmes qui l'opèrent.

Depuis Platon l'homme s'interroge sur ses émotions. Succédant à la tradition philosophique, les sciences modernes se sont emparées du sujet. Il est, aujourd'hui encore, plus facile de définir les émotions par ce qu'elles *font*, plutôt que par ce qu'elles *sont* réellement.

Dans le premier chapitre, nous essayons de cerner ce qu'est une émotion. Nous tâchons de comprendre sur quelles bases théoriques se construisent les différents modèles de l'émotion.

Nous présentons ces 3 principaux modèles théoriques et nous les comparons.

Le premier, issu des travaux de Charles Darwin au XIX$^{\text{ème}}$ siècle, pose l'existence de 6 émotions de base, programmées génétiquement comme des réponses de survie aux événements de l'existence. Au cours du XX$^{\text{ème}}$ siècle, deux théories alternatives émergeront : les théories multi-dimensionnelles et de l'évaluation cognitive.

De ces 3 courants théoriques ressortent des points de consensus, mais également de divergence, qui nous aideront par la suite à comprendre comment elles ont inspiré la technologie informatique de reconnaissance faciale de l'émotion.

Dans le second chapitre, nous abordons la construction technique des systèmes automatiques d'analyse de l'émotion basés sur la reconnaissance faciale, et nous cherchons à expliquer comment s'établit l'inférence d'une notion abstraite d'émotion à partie des signes visibles qu'affichent les mouvements d'un visage.

Nous commençons par une distinction entre reconnaissance de l'émotion et analyse de l'activité musculaire faciale. Ce chapitre illustre la très grande dépendance des systèmes de détection à la seule théorie des émotions de base de Paul Ekman. C'est le modèle dominant et le plus largement utilisé par les concepteurs de systèmes informatiques.

En établissant une liste de 6 états émotionnels discrets et en les liant à des combinaisons d'action du système musculaire du visage, les travaux de Paul Ekman apportent aux sciences informatiques une représentation de l'émotion modélisable en informatique et accessible aux techniques de machine learning.

En machine learning, quel que soient le modèle ou les techniques dont on dispose pour classer les données, aucun résultat fiable ne peut sortir d'un classifieur sans de larges quantités de données expérimentales pour les entraîner. Le troisième chapitre aborde la collection de données pour l'entraînement des classifieurs.

Constituer une base expérimentale est un travail essentiel, qui déjà repose sur des choix subjectifs dans sa représentation : le caractère posé ou spontané des images, leur développement temporel ou la manière de les codifier sont autant de choix qui influent sur la fiabilité de l'inférence que l'on fait à partir d'une expression.

Nous proposons une sélection de critères sur lesquels serait basé un *dataset* idéal.

Nous illustrons également de quelle manière la collecte et la codification, jusqu'ici manuelles, de grandes quantités d'image mobilise d'importantes quantités de ressources, et en quoi cela

représente un obstacle au passage d'un environnement de test à un fonctionnement dans les conditions de la vie de tous les jours.

Nous exposons les solutions envisagées pour automatiser cette partie du travail.

Dans le quatrième chapitre, nous avons souhaité évoquer un concept alternatif à la théorie des émotions de base qui domine le champ de l'analyse de l'émotion.

Le modèle de l'évaluation cognitive, ou modèle de l'*appraisal*, est ici présenté pour une reconnaissance de l'émotion fondée non plus sur des *patterns* génétiques, mais sur une séquence d'évaluations par l'individu, dans son environnement, sa culture et son contexte.

Bien que minoritaire en reconnaissance de l'émotion, le modèle de l'*appraisal* remet en cause l'aspect inné des émotions de base et prétend à la reconnaissance d'un éventail d'états plus large et plus subtils.

En conclusion de ce travail, nous nous interrogeons sur la validité des fondements théoriques et méthodologiques de l'approche dominante par les émotions de base. La très grande dépendance de la plupart des systèmes vis-à-vis des concepts théoriques portés par Paul Ekman s'explique, comme nous l'exposons au deuxième chapitre, par une approche pragmatique, mais aussi par la visibilité académique et médiatique de son travail. Nous questionnons en premier lieu la validité scientifique du modèle des émotions de base et de sa capacité à saisir la finesse de la sémantique d'une émotions. Nous adressons également la question du régime de justice que pose la généralisation des techniques de reconnaissance faciale du point de vue sociétal. Qu'en est-il de la neutralité de cette technologie ? Enfin, nous soulignons comment cette technologie est potentiellement porteuse de biais induits par la construction des bases de données et la fragilité du cadre d'expérimentation actuel pour entraîner les modèles.

L'approche que nous tentons s'est voulue interdisciplinaire dès le départ, ce qui aura nécessité l'exploration d'une bibliographie élargie aux sciences humaines et à la psychologie notamment. Nous pensons que le travail livré propose une entrée en matière accessible à un public plus large que celui des informaticiens et que la sélection des sources proposée fournira à tout lecteur une base intéressante pour approfondir le sujet.

Chapitre 1

Les émotions

Dans ce premier chapitre, nous introduisons le concept d'émotion.
Nous tentons d'abord d'en donner une définition, nous passons ensuite en revue les principaux courants de la théorie des émotions, nous en détaillons la fonction et la structure, et enfin nous nous attachons à établir la nature motivationnelle des émotions.

Les questions auxquelles nous nous proposons de répondre au terme de ce premier chapitre sont : "Qu'entend-on par émotion ?" et "de quoi sont-elles porteuses comme indications sur le sujet qui les affiche" ?

1.1 Définitions

Autrefois considérées par les philosophes, de Platon à Descartes, comme radicalement opposées à la Raison, les Passions, telles qu'on nommait les émotions à l'époque, ne pouvaient se concevoir que comme venant perturber la pensée noble, désirable, élévatrice, qui distingue l'animal de l'être humain, l'homme érudit de la femme ou de l'enfant.

On peut considérer que les années 60 marquent un changement dans la psychologie sociale du traitement des émotions. Elles sont reconnues fondamentales aux rapports entre être humains et à la cohésion sociale.
Les travaux menés dans la lignée de Charles Darwin contribuent ainsi à mettre en lumière le rôle adaptatif des émotions, ainsi que la double fonction des comportements sociaux, agissant dans un premier temps comme stimuli des émotions, pour ensuite être en retour modelés par ces dernières.

En dépit de ces avancées, on peut considérer qu'il n'existe pas à ce jour de définition précise de l'émotion.
« Il est intéressant de noter que la plupart des définitions font référence à leurs caractéristiques, (...),leurs rôles ou leur fonctions (...). Le fait de les décrire si souvent par leur caractéristiques plutôt que par leur nature essentielle souligne la mauvaise compréhension que nous avons encore de ces phénomènes complexes [Hudlicka, 2015] ».

Selon Nico Frijda, « il n'existe donc pas de véritable théorie des émotions. Par véritable théorie de l'émotion j'entends une théorie de l'organisme humain, ou des systèmes biologiques en général, au sein desquels l'émotion a sa place propre, parmi d'autres composantes comme le traitement de l'information et de l'adaptation [Frijda, 1989] ».

Néanmoins, l'approche actuelle des émotions permet de dégager un certain nombre de points sur lesquels un consensus se dégage, tels que la décomposition des émotions en plusieurs composantes, leur nature hautement adaptative et leur fonction sociale.

Ainsi, la définition suivante en est proposée dans [Nugier, 2009] : « phénomène multi-composant adaptatif (...) pouvant être caractérisé par les expressions réactives, comme le sourire, le froncement de sourcil, l'intonation de la voix, la posture ; par les réactions physiologiques comme la fréquence cardiaque, le flux sanguin, la production des larmes ; par les tendances à l'action et les réactions comportementales, comme l'attaque, l'évitement, la fuite, la recherche de support social ; par les évaluations cognitives, comme la pensée d'avoir été injustement traité par une autre personne ; et enfin par l'expérience subjective (ou sentiments subjectifs), c'est-à-dire ce que l'on pense ou dit ressentir ».

1.2 Les théories des émotions

1.2.1 Une histoire de la théorie des émotions

Les premières études sur ce qu'est l'émotion coïncident avec les débuts de la psychologie en tant que discipline de recherche indépendante, et plus particulièrement avec les travaux de Charles Darwin à la fin du XIX$^{\text{ème}}$ siècle.

Darwin sera le premier à poser un corpus théorique sur les émotions, qu'il définit en 7 thèses, qui posent les bases de sa théorie évolutionniste : les émotions sont innées, adaptatives et universelles.

William James [James, 1884], plus tard, renverse, en un sens, la logique darwinienne, arguant que ce n'est pas la naissance d'un état mental émotionnel qui conditionne les réactions corporelles, mais l'inverse.

C'est parce que la vue d'un ours déclenche le tremblement et l'action de fuir que nous ressentons ensuite la peur. Il existe bien un ensemble de réflexes innés, mais c'est par la prise de conscience de ces réflexes que naît l'émotion.

Le début du XX$^{\text{ème}}$ siècle marque le retour à une forme de *behaviorisme* dur. Seul le comportement directement observable peut faire l'objet de l'étude des sciences du comportement. Les émotions sont réduites à des réactions physiologiques dénuées de tout caractère adaptatif. Les émotions ne représentent qu'un état de désordre mental dont il convient de s'affranchir.

Les travaux de Stanley Schachter [Schachter and Singer, 1962] ouvrent l'étude de l'émotion à l'aspect cognitif. Ce sont les débuts d'une approche proprement psychologique à l'étude des émotions.

Il propose que l'émergence d'une émotion va de pair avec une interprétation par le sujet de la situation, et de son contexte.

Les émotions on un caractère diffus. En conséquence, l'individu cherchera constamment à balayer son environnement pour y chercher des explications cognitives.

Une émotion résulte au final de la conjonction d'une activation physiologique et d'une évaluation cognitive.

La théorie de l'évaluation cognitive sera par la suite largement développée et raffinée, par les travaux de Magda Arnold [Arnold, 1960], qui définit le concept d'*appraisal* qui théorise qu'une émotion est toujours « colorée »par la relation propre qu'a l'individu à l'objet ou au stimulus ; sa culture, sa mémoire, ses croyances.

Klaus Scherer [Scherer, 2001] développe le modèle et ouvre la voie à une représentation de l'émotion qu'il veut plus compatible avec la modélisation informatique.

Il propose 5 niveaux à travers lesquels une situation particulière est analysée : nouveauté, plaisir, pertinence, capacité à faire face à la situation (*coping*), compatibilité avec les normes.

L'ensemble des variables de ces 5 composants déterminera formellement la qualification de l'émotion.

Si les représentations peuvent diverger, on trouve en point commun des différentes théories de l'*appraisal* un processus d'évaluation permanent, et la confrontation d'événements à la représentation que l'individu a du monde et des normes, et de ses objectifs.

Enfin, dans les années 60, émerge le courant néo-darwiniste, reprenant les conceptions adaptatives et universelles de Darwin.

Pour survivre, l'organisme a mis en place des mécanismes d'amplification des stimuli qui commandent à l'organisme de passer à l'action.

Silvan Tomkins [Tomkins, 1962] définit 9 programmes innés de l'affect, qui, une fois activés, formulent des réponses, dont le médium premier de différentiation est l'expression du visage. Les travaux de Paul Ekman et Carroll Izard [Izard, 2007] poursuivent cette logique.

A ce jour, on peut considérer ces deux dernières approches comme les plus représentées dans les travaux de recherche sur l'émotion, bien que plus complémentaires qu'opposées, car elles ne s'intéressent pas tout-à-fait aux mêmes facteurs [Rimé, 2009].

Dans la suite, nous proposons de décrire plus en détail les 3 modèles les plus utilisés, ouvrant la porte au traitement de l'émotion dans un système d'information : théories évolutionniste, dimensionnelle et à composants.

1.2.2 Théories évolutionnistes

Charles Darwin pose en 1872 dans *L'expression des émotions chez les hommes et les animaux* les premiers jalons d'une théorie des émotions.

Il pose deux points capitaux : les émotions sont universelles et adaptatives, en ce sens qu'elles contribuent à la perpétuation de l'espèce en lui permettant de répondre de manière efficace aux menaces de son environnement. Ces émotions, dites de base, (ou primaires, ou discrètes), innées, viennent en soutien de sa théorie de l'évolution. Ainsi par exemple, le dégoût serait associé à une geste d'expulsion de la nourriture et de pincement du nez pour éviter les odeurs.

Cette approche, largement utilisée aujourd'hui, dépasse les constatations de Charles Darwin : l'émotion est vue comme un « processus organisateur à haut niveau »[1], ou comme le système motivationnel primaire du comportement humain dans la théorie des émotions différentielles de Caroll Izard [Izard, 2009].

Le point central dans la théorie évolutionniste est l'identification de 6 émotions de base (joie, colère, tristesse, dégoût, surprise, peur), en plus des émotions complexes, résultant de combinaisons d'émotions basiques.

L'existence de ces émotions universelles postule dès lors d'une part qu'il existerait des stimuli universels (le décès d'un être cher serait une condition universelle de la tristesse), et d'autre part, qu'il existe pour ces émotions de base, des patterns neuronaux spécifiques. Nous revenons par la suite sur ce point.

1.2.3 Théories bi-dimensionnelles

A l'origine, ce courant se fonde sur la théorie de Wundt [Wundt, 1897], qui identifie trois dimensions de base pour définir le sentiment subjectif de l'émotion : caractère excitant/déprimant, plaisant/déplaisant, tension/relaxation.

Depuis, les tenants contemporains de cette approche ont ramené à deux le nombre de dimensions qualifiant l'émotion.

S'ils sont d'accord avec le fait que les émotions sont le fait de l'évolution, ils ne considèrent pas pertinent de les réduire à une liste discrète d'émotions basiques. Considérer la peur ou la colère comme catégories émotionnelles de base « représente maintenant un obstacle majeur à la compréhension de ce que sont les émotions et de comment elles fonctionnent [Barrett, 2006] ».

Selon cette approche, toute émotion peut toujours être définie en recourant à des dimensions mesurables et indépendantes, telles que les manifestations physiologiques. Le modèle le plus couramment utilisé fait usage des dimensions de valence et activation, également considérées comme universelles.

La méthodologie de définition habituelle consiste à questionner des individus sur leur sentiment d'être plus ou moins stimulé, et sur quelle échelle entre plaisir et déplaisir. Cette méthode est considérée comme fiable et propice au traitement statistique. En revanche, le résultat ne dit rien des stimuli qui ont conduit à susciter l'émotion, ni comment l'évaluation en a été faite par l'individu.

James Russel [Russell and Pratt, 1980] développe un modèle proposant les 2 dimensions de *valence* (plaisir/déplaisir) et *activation* (faible/forte).

Le modèle circulaire est appelé *circumplex*, et postule que chaque état affectif peut en permanence être représenté par une valeur plus ou moins élevée sur chacun des deux axes.

La principale critique formulée à l'encontre de ce modèle [Coppin and Sander, 2010] est qu'il est difficile de distinguer des émotions telles que la peur et la colère, toutes deux caractérisées par des niveaux proches de valence et d'activation.

1. Cosmides, Leda, and John Tooby. "Evolutionary psychology and the emotions." Handbook of emotions 2 (2000) : 91-115.

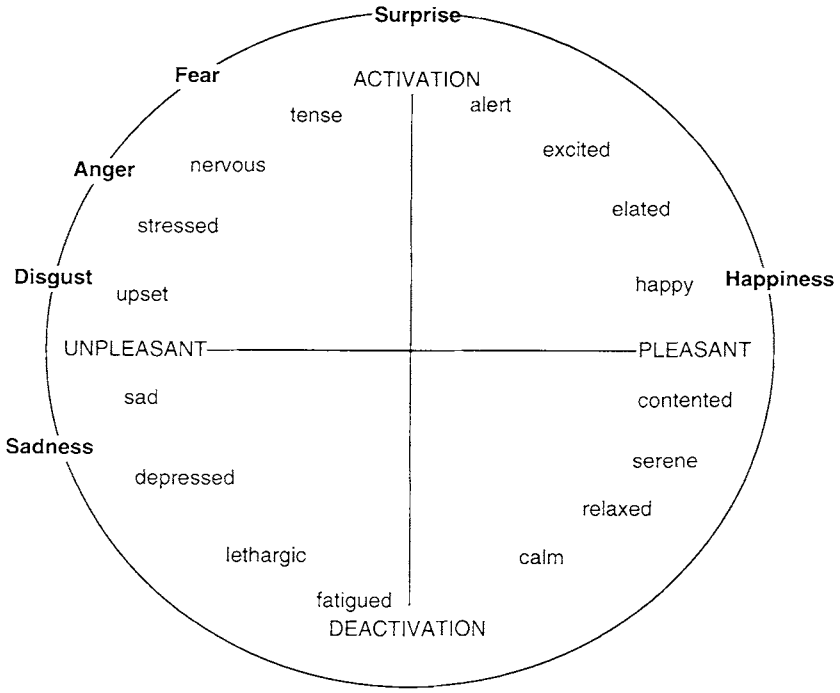


FIGURE 1.2.1 – Le modèle bi-dimensionnel de Russell.

Source : [Russell and Barrett, 1999]

Comme nous le verrons dans la suite, ceci ne manquera pas de poser problème lorsqu'il s'agira de choisir une représentation qui puisse être évaluée avec précision par un système informatique. D'autre part, il n'y a pas consensus sur le nombre et la nature des dimensions à utiliser.

Albert Mehrabian [Mehrabian, 1996] y ajoute la dimension de dominance, quand Elizabeth Duffy [Duffy, 1941] résume l'émotion à l'évaluation d'un niveau d'énergie. Pour [Coppin and Sander, 2010], le problème majeur de cette approche réside dans le manque de consensus sur les concepts auxquels il est fait appel.

Le point de vue le plus souvent partagé fait appel aux valeurs de valence/activation, ne disant rien ce faisant de l'intensité de l'émotion. L'alternative est d'y adjoindre l'intensité comme 3$^{\text{ème}}$ dimension, ceci posant le problème de savoir si elle porte sur la valence ? l'activation ? les deux ?

1.2.4 Théories de l'évaluation cognitive *(appraisal)*

Les théories (car il en existe plusieurs courants, que nous ne détaillons pas tous ici) de l'évaluation cognitive prennent leur source dans les travaux de Magda Arnold [Arnold, 1960] et Richard Lazarus [Lazarus, 1966] , qui théorisent le concept d'*appraisal* comme explication au

fait qu'un même stimulus peut être évalué différemment par différents individus à des moments différents.

Pour les défenseurs de la théorie de l'évaluation cognitive, le ressenti émotionnel est déterminé par la signification personnelle que nous donnons aux événements. Dans cette perspective, les individus explorent l'environnement en continu, réagissant aux stimuli à travers un processus d'évaluation cognitive nommé appraisal, fondé sur un ensemble de critères définis.

Les composantes les plus communément admises sont la pertinence, l'implication, le potentiel de maîtrise de la situation et l'évaluation normative [Grandjean and Scherer, 2009], mais on cite également la nouveauté, le caractère prédictible ainsi que l'importance pour le but à atteindre. C'est de l'évaluation séquentielle de ces différentes composantes que naît le sentiment émotionnel.

L'idée centrale est qu'à chaque structure d'évaluation différente correspond une émotion, décrite sous forme de pattern d'évaluation cognitive.

Parmi les principaux tenants de la théorie de l'appraisal, citons Nico Frijda, qui lie la naissance d'une émotion au caractère pertinent des événements, soit «quand ils touchent à un ou plusieurs des préoccupations (*concerns*) du sujet», ("concern" au sens d' « état d'esprit d'une personne qui s'intéresse particulièrement à quelqu'un, quelque chose, s'adonne à une activité, veut parvenir à un but [Frijda, 1986] »).
Nico Frijda théorise le processus d'évaluation de la pertinence des situations comme débouchant sur une tendance à l'action, principale caractéristique à ses yeux, permettant à l'individu de faire face à une situation [Frijda, 1986].

Pour Klaus Scherer, auteur du Component Process Model (CPM), l'évaluation suit une séquence de Stimulus Evaluation Check (SEC), qui contribue à l'élicitation une émotion. Il appréhende cette genèse comme un processus dynamique, parlant plutôt d'« épisode »plutôt que d'« état »émotionnel [Frijda, 1989].
Contrairement aux modèles précédents, le modèle par composants ne se réfère pas qu'à un seul aspect : l'émotion est en effet multi-dimensionnelle et possède 5 composantes :

- une composante d'évaluation cognitive des stimuli,
- une composante physiologique,
- une composante d'expression motrice de l'expression faciale, de la posture corporelle et du geste,
- une composante motivationnelle induisant une tendance à l'action,
- une composante subjective de ressenti de l'émotion, sorte de résultante des changements s'opérant dans les autres composantes [Scherer, 2001].

La figure 1.2.2 propose une description du modèle des processus composants : le process est représenté comme une suite séquentielle d'objectifs d'évaluation cognitive (séquence de gauche à droite). Chaque évaluation influe (flèches noires descendantes) sur chacune des composants émotionnels(physiologique, tendance à l'action, motrice,...), ainsi qu'avec les autres fonctions cognitives (attention, mémoire, raisonnement,...).

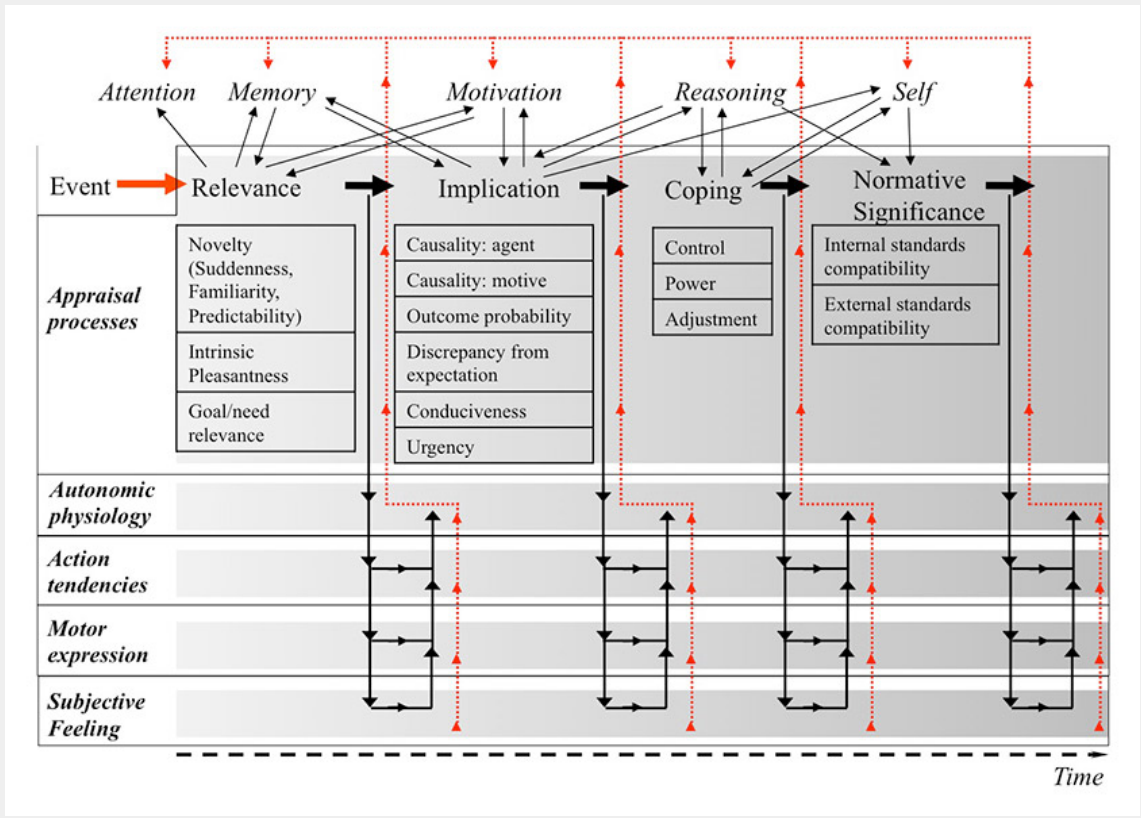


FIGURE 1.2.2 – Le modèle de processus par composants de Scherer

Source : [Brosch et al., 2013]

Enfin, ces changements dans les différents composants influent également (flèches rouges ascendantes) sur les fonctions cognitives, et récursivement sur les critères guidant l'appraisal.

Les critiques envers la théorie de l'appraisal sont de deux types ; premièrement, le caractère cognitif de l'expression émotionnelle est contesté.

Par exemple [Berkowitz, 1993], qui s'intéresse particulièrement à l'émergence de la colère et de la frustration, souligne les ambiguïtés dans la vision cognitiviste, ou bien [Zajonc, 1984], pour qui il n'existe pas de preuve empirique de la théorie de l'appraisal, au contraire, défendant l'idée qu'il existe des affects indépendants et antécédents au processus cognitif.

Ensuite, les troubles émotionnels ne sont pas pris en compte. Par exemple : « le fait, pour un phobique des araignées, d'acquérir la connaissance explicite du fait qu'une araignée est sans danger ne l'empêchera pas d'en avoir peur » [Griffiths, 2004].

En résumé, la figure 1.2.3 illustre de manière schématique les processus d'apparition de l'émotion dans les 3 grands courants.

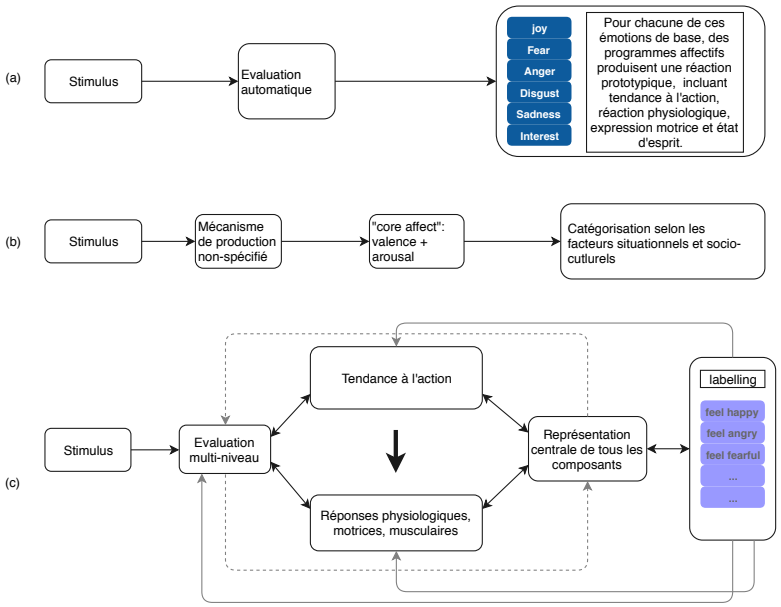


FIGURE 1.2.3 – Comparaison des modèles d'émotion (a) évolutionniste, (b) dimensionnel, (c) d'évaluation cognitive.

Source : tiré de Scherer, K. R. (2009). Emotions are emergent processes : they require a dynamic computational architecture. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1535), 3459-3474.

1.3 Émotion et tendance à l'action

Puisqu'il existe un consensus sur la nature adaptative de l'émotion, il est logique de poursuivre la réflexion sur le lien entre émotion et prise de décision, adaptation comportementale et passage à l'action.

Afin d'établir l'existence de ce lien, nous passons en revue ce qui apparaît comme convergent dans trois approches différentes mais complémentaires, que sont la philosophie, la psychologie et les neuro-sciences.

1.3.1 Contribution de la philosophie

La dualité entre émotion et prise de décision, ou entre émotion et Raison questionne les philosophes depuis l'Antiquité.

Ainsi, pour Descartes, le clivage net entre émotion et raison empêche toute affect ou ressenti émotionnel de prendre part au processus noble de la prise de décision rationnelle.

Dans « L'erreur de Descartes »[Damasio, 2006], Antonio Damasio prend le contrepoint de la théorie cartésienne en vogue à l'époque pour établir au contraire les inter-relations fortes entre le cerveau et les émotions dans la relation des individus à l'environnement.

Par l'entremise de la neuropathologie, il établit le rôle des émotions (innées et acquises) à anticiper le futur et à remodeler les comportements à la lumière des intérêts propres ou objectifs de survie.

Dans « Spinoza avait raison »[Damasio, 2003], il fait le lien entre les neuro-sciences et la pensée de Spinoza : les sentiments expriment la lutte de l'individu pour accéder à l'équilibre, corps et esprit se fondent et aspirent au bien-être par les émotions et les sentiments.

Pour Hume dans "A Treatise of Human Nature", la raison est, et doit rester subordonnée aux passions, et ne peut prétendre à rien d'autre qu'à leur obéir[Hume, 1978].

1.3.2 Approche par les neuro-sciences

L'approche par les neuro-sciences cognitives permet d'établir les liens entre l'émotion et le rôle de l'amygdale, le cortex pré-frontal et l'axe hypotalamo-hypophyso-surrénalien (HHS).

L'amygdale tout d'abord, dont les dysfonctionnements mettent en lumière les troubles d'ordre émotionnel et décisionnel qui en résultent [Bechara, 2004].

Le cortex pré-frontal, comme centre de l'évaluation cognitive et de planification de l'action [Salzman and Fusi, 2010], et l'axe HHS, composante fondamentale de l'émotion, organe principal du stress [Dickerson and Kemeny, 2004].

Enfin, Hagar Goldberg *et.al*, dans [Goldberg et al., 2014], établit par l'étude neuro-biologique des informations dans la voie dorsale du cortex, l'association entre émotion et préparation à l'action.

1.3.3 Données de la psychologie

On peut recenser aujourd'hui un nombre significatif de contributions dans la recherche en psychologie et en sciences du comportement, pour étayer l'idée que « *les émotions déterminent comment nous percevons le monde, comment nous le mémorisons et quelles décisions nous prenons* »[2].

La dualité conflictuelle entre raison et émotion est désormais démentie au profit d'une approche de l'émotion comme un guide nous permettant d'appréhender les situations complexes de l'environnement.

Les travaux de Jennifer Lerner *et.al* [Lerner et al., 2015] offrent une compilation de 35 ans d'efforts de recherche autour de la relation entre émotion et prise de décision.

Intégrer les contraintes cognitives et situationnelles dans le processus de décision implique dès lors d'identifier le rôle que peuvent y jouer les émotions. C'est le point de départ d'un effort croissant de recherche multidisciplinaires sur l'émotion comme mécanisme de décision [Lerner et al., 2015].

L'émotion est désormais reconnue comme facteur majeur de la prise de décision par Paul Ekman [Ekman, 2007], Nico Frijda [Frijda, 1988], Richard Lazarus [Lazarus, 1991], Klaus Scherer & Ekman [Scherer and Ekman, 2014], parmi les contributeurs les plus connus.

Il y a un consensus large considérant la prise de décision comme un canal à travers lequel nos émotions tendent, même inconsciemment, à limiter les affects négatifs (tristesse, culpabilité), et à maximiser les affects positifs (fierté, joie) [Lerner et al., 2015].

Des travaux recensés, ciblant essentiellement les *behavioral studies*, 8 thèmes de recherche sont dégagés, qui mettent en lumière l'influence des émotions sur la prise de décision.

Par exemple, la valence n'est qu'une des dimensions parmi d'autres qui incitent à la prise de décision. Lerner *et al.* proposent l'Appraisal Tendency Framework (ATF) [Lerner and Keltner, 2000], [Lerner and Keltner, 2001], un modèle qui, poussant plus loin que la seule échelle de valence, associe le process d'*appraisal* complet d'une émotion particulière aux choix qu'il induit.

Le postulat repose sur la triple hypothèse que

a) à chaque ensemble de dimensions cognitives correspond une manifestation émotionnelle,

b) les émotions jouent un rôle de chef d'orchestre, associant les réponses (comportementales, physiologiques,...) nécessaires à l'individu pour interagir avec les événements,

c) les émotions ont un caractère motivationnel lié tant à leur intensité qu'à leur nature qualitative.

Elle sont en soi porteuses de "tendance à l'action" [Frijda, 1986], ou d'objectifs implicites qui sous-tendent la réponse la plus adaptative aux sollicitions de l'environnement.

De la même manière que les émotions sont, pour Frijda, porteuses de tendance à l'action, l'ATF pose la prédisposition de l'individu à procéder au mécanisme d'*appraisal* de manière spécifique.

Un autre axe de recherche conclut que les émotions modèlent les décisions par l'entremise de l'activation de buts. Frijda *et.al* [Frijda, 1986] associent par exemple la colère au besoin de

2. Brosch, T., Scherer, K. R., Grandjean, D., Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. Swiss Medical Weekly, 143, w13786

confrontation physique avec un individu ou un obstacle. En parallèle, la préparation à l'« état d'action »se manifeste par l'afflux de sang dans la tête ou dans dans les mains.

Ces tendances à l'action correspondent à des *patterns* d'évaluation cognitive. Ainsi par exemple, le sentiment d'anxiété, impliquant une menace incertaine, va de pair avec une tendance à l'action visant à réduire le niveau d'incertitude [Lazarus, 1991].

De la même manière, dans [Raghunathan and Pham, 1999], l'expérimentation vise à confronter les effets des états de tristesse et d'anxiété à des comportements liés au jeu ou à la sélection d'une profession, concluant que la tristesse favorise les choix de type haut niveau de risque/récompense, quand l'anxiété porte plutôt vers des options de risque/récompense faible.

Un dernier thème établit l'influence des émotions sur les décisions inter-personnelles. A ce jour, les recherches sur ce thème permettent de conclure au rôle joué par les émotions pour :

a) aider les individus à comprendre, reconnaître leurs intentions réciproques,

b) encourager tel ou tel comportement chez autrui,

c) évoquer chez autrui le partage ou la réciprocité des émotions.

Pour terminer, Caroll Izard, défenseur de la thèse évolutionniste des émotions, évalue lui aussi l'influence des émotions à altérer le comportement.

Il introduit une distinction entre les émotions de base et les schémas d'émotions (interaction entre émotion et cognition suggestives de ressentis ou comportements, allant d'épisode momentané jusqu'à la définition de traits fondamentaux).

Le ressenti d'une émotion de base affecte certainement l'action, mais pas l'évaluation de haut niveau, tandis que le ressenti d'un schéma d'émotion affecte tant l'action que le processus cognitif [Izard, 2008].

1.4 L'expression faciale des émotions

Les canaux d'expression d'un ressenti émotionnel peuvent être multimodaux (expression faciale, ton de voix, gestuelle, mouvement du corps). Dans la suite de ce mémoire, nous nous focalisons désormais sur son expression faciale.

Après avoir parcouru la construction théorique des émotions, nous nous attachons dans cette section à comprendre ce qu'elles sont en mesure de dire de celui qui les affiche.

La question que nous posons dans cette section est de savoir, une fois posée la possibilité de capter l'expression d'un visage, ce qu'elle est susceptible d'exprimer au sujet de son émetteur, dans son contexte personnel.

1.4.1 Universalité des émotions

La question posée par le caractère universel ou socio-culturel des émotions renvoie à une controverse longue de plusieurs décennies entre partisans des écoles *darwiniste* et *behavioriste*.

Le premier à poser le postulat d'émotions universellement partagées est Charles Darwin. Pour

lui, les expressions émotionnelles renvoient à des comportements rudimentaires développées par les espèces pour assurer leur subsistance ou leur survie. Les émotions ont donc la triple fonction d'adaptation, de communication d'un état affectif, et de manifestation des stimuli par le système nerveux.

Il en découlerait que les émotions de base sont donc partagées par tous les individus, indépendamment de leur sexe, age, culture ou ethnie, selon les résultats des études menées en psychologie dans les années 60 et 70.

Dans ces expérimentations, des observateurs issus d'Amérique, d'Europe, d'Asie et d'Afrique sont amenés à apparier des expressions faciales avec une liste discrète de 6 émotions.

Il en résulte, dans tous les cas, un taux d'accord inter-juges supérieur au hasard, y compris dans les expériences avec des civilisations isolées, n'ayant pas pu faire l'apprentissage de nos conventions sociales.

S'ensuivit néanmoins une intense controverse sur la méthodologie utilisée, et sur la nature « exagérée »des expressions sur les images posées. Nous revenons par ailleurs sur ce point, toujours pas tranché, dans la dernière section de ce travail.

Plus récemment, Haidt et Kelner [Haidt and Keltner, 1999] et Frank et Stennett [Frank and Stennett, 2001], ont apporté des preuves supplémentaires en faveur de la thèse de l'universalité.

En procédant par une méthode de choix forcé améliorée, ils répondent aux critiques relatives à la méthodologie, et étendent leurs conclusions à d'autres expressions que les 6 émotions discrètes, telles que la honte, l'embarras, le mépris.

La thèse de l'universalité est cependant nuancée par les mêmes Haidt et Kelner, qui pointent également l'influence de :

a) l'apprentissage culturel : Les taux d'accord restent significativement plus élevés entre juges de même culture,

b) les « règles d'expression »(*expression rules*) qui peuvent avoir tendance, dans certaines cultures, à limiter l'intensité d'une expression faciale en fonction du contexte social. (Ce qui est également la conclusion d'une autre étude menée par Biehl et Ekman impliquant des sujets américains et japonais Biehl et al. [1997]).

En conclusion, s'il existe un consensus large quant à la faculté à reconnaître et décoder une émotion à travers des cultures différentes, il manque encore de fondements théoriques et d'expérimentations pour affirmer qu'il en va de même pour l'« encodage »(par le biais de son expression faciale) au sein de contextes sociaux ou culturels différents ».

1.4.2 De quelles informations les expressions faciales sont-elles porteuses ?

Ici encore, on constate une divergence de vue entre la conception évolutionniste, défendue par Paul Ekman, qui postule que l'expression faciale renvoie de l'information sur l'état interne réel de l'individu (c'est l'approche émotion-expression), et la conception de l'écologie comportementale, pour laquelle au contraire, l'émotion est à voir principalement comme vecteur de communication par l'individu de ses comportements sociaux et intentions.

Parkinson [Parkinson, 2005] compare les deux approches à travers les travaux d'Ekman [Ekman, 1971] et Fridlund [Fridlund, 1994].

Ekman postule que les mouvements du visage, tirés à l'origine de la théorie de l'évolution, peuvent avoir évolué depuis les origines, précisément pour exprimer des émotions, plutôt que des préoccupations triviales de survie.

Il existerait des programmes neuro-moteurs innés, liant les expressions de base aux composantes musculaires du visage. C'est ce qu'il définira comme les *Facial Action Units*, responsables de commander les muscles du visage.

Si le contexte social nous enseigne les causes pouvant amener à une émotion, ou comment les « *display rules* »nous recommandent de les atténuer, c'est bien un programme inné qui fait le lien entre l'émotion élicitée et sa transcription en signaux perceptibles dans le visage. Il est fait référence à cette théorie sous l'appellation « *neuro-cultural theory* ».

Alan Fridlund, par opposition, préconise que les expressions faciales sont

a) un vecteur de *communication* vers les autres plutôt que d'*expression*,

b) l'annonce de « motivations sociales » ou d'« intentions comportementales »[Parkinson, 2005].

La fonction adaptative n'a de sens pour lui que si elle permet aux congénères d'ajuster leur comportement aux intentions affichées sans ambiguïté.

S'il ne semble pas possible d'émettre un avis tranché entre ces 2 conceptions, il ressort les conclusions suivantes de la recherche sur le sujet :

- Les émotions ne peuvent se justifier comme seules déclencheuses des expressions faciales qu'elles suscitent.
- Un individu peut, lors d'un épisode émotionnel, ne pas marquer d'expression faciale, même hors l'influence de tierces personnes ou de « display rules ».
- Un grand nombre de manifestations émotionnelles sont induites par le contexte social.
- Il n'existe pas de démonstration d'expression faciale liée à une émotion particulière, de manière purement et strictement non ambiguë.

En conclusion, il est fait la recommandation d'axer la recherche sur ce thème vers l'analyse en temps réel des actions faciales, tant comme conséquences d'*appraisals* pour le sujet, que comme conséquences de comportements en réaction chez les congénères.

1.4.3 L'hypothèse de rétroaction faciale

Pour en terminer avec cette section, il nous semble enfin utile, pour compléter la compréhension de ce qu'est une émotion, d'invoquer l'hypothèse de rétroaction faciale.

Cette hypothèse pose que la fonction de l'expression d'une émotion ne se limite pas à communiquer sur un état interne et des tendances à l'action, mais que cette expression (respectivement son inhibition) agit également en retour sur l'émotion elle-même, la rendant plus intense (respectivement plus modérée).

Cette théorie est formulée sous la forme de 2 hypothèses :

a) *l'hypothèse de modulation* : les expressions faciales jouent sur l'intensité du ressenti émotionnel (par exemple : l'homme qui accompagne sa colère de gestes violents augmente l'intensité de sa colère).

b) *l'hypothèse d'induction* : l'acte de simuler ou imiter une émotion tendra à susciter l'expérience de cette émotion, même en l'absence de stimuli. (par exemple : simuler un sourire est de nature à faire émerger un sentiment de joie.

Paula Niedenthal *et.al* [Niedenthal et al., 2009] présentent différentes méthodologies d'expérimentation ayant mené à ces conclusions.

Typiquement, des participants sont amenés à simuler des expressions faciales de manière soit à exagérer, soit à inhiber une réaction émotionnelle à des stimuli (chocs électriques, images amusantes ou tristes, odeurs, manifestations festives), indépendamment de l'intensité réelle de ces stimuli.

Les résultats illustrent l'influence qu'ont les expressions faciales à amplifier ou atténuer l'intensité de l'expression ressentie par le sujet.

1.5 Conclusion

Les épisodes émotionnels qui peuvent nous paraître simples tant nous en rencontrons chaque jour de notre vie sont en réalité complexes et de larges pans en sont toujours incompris. Il y a une vraie difficulté littéraire pour un informaticien à aborder les concepts du domaine et à tenter une synthèse à travers les disciplines. Néanmoins nous pensons que comprendre la théorie que posent les sciences affectives permet à l'informaticien de poser des choix plus judicieux sur la manière de la représenter ensuite.

Les émotions peuvent être vues comme phénomènes adaptatifs innés dans la tradition darwinnienne (Ekman, Izard, Tomkins), définies sur une schéma bi-dimensionnel représentant les 2 axes de valence et activation (Russell), ou enfin comme le résultat d'un processus d'évaluation (*appraisal*) par l'individu, coloré par les savoirs, le contexte et les expériences antérieures (Arnold, Scherer).

Le caractère universel (du moins pour les émotions de base) est largement partagé. Avec les nuances que le contexte culturel peut en altérer l'expression(« display rules »), ou le décodage. Il existe des différences résultant de l'apprentissage culturel entre cultures différentes. Cela a été mis en avant notamment par comparaison entre individus de culture plus individualiste (culture occidentale), ou collectiviste (Japon).

Il existe donc un socle d'émotions universellement identifiables par l'humain, par delà les différences culturelles.

Il n'existe pas d'opposition irréconciliable entre le raisonnement rationnel d'une part, et les émotions d'autre part. Il ressort des recherches menées au cours des 35 dernières années, qu'il existe un lien fort entre les émotions d'une part, et la prise de décision, la tendance à l'action, ou l'expression d'intentions.

L'expression faciale des émotions a pour fonctions de communiquer de l'information sur l'état affectif interne, mais également d'informer les autres individus de ses intentions.

Enfin, selon la théorie de la rétroaction faciale, le seul fait de simuler, ou d'inhiber l'expression d'une émotion suffit à induire ou moduler l'intensité du ressenti de cette émotion par le sujet.

Pour terminer cette section, nous présentons (voir Table 1.1) un récapitulatif des 3 grands courants de la théorie des émotions.

	Théories évolutionnistes	**Théories bi-dimensionnelles**	**Théories des composants**
Origines théoriques	C. Darwin	W. Wundt	M. Arnold
Principales contributions	Ekman, Frijda, Izard, Tomkins	Russell, Mehrabian, Feldman, Barrett	Lazarus, Scherer
Sémantique de l'émotion	Emotions de base : nombre limité de patterns innés conditionnant l'individu à répondre de manière adaptative à son environnement	Axe multi-dimensionnel (valence, arousal, dominance)	Appraisal : l'émotion découle de l'évaluation en parallèle de variables d'appraisal. Les appraisals successifs forment les composantes de l'émotion.
Formalisation de l'émotion	Set limité d'émotions discrètes (6 émotions de base)	Un set plus étendu, par la combinaison de 2 ou 3 coordonnées	Le set le plus large, par 5 composantes dérivées de l'ensemble des variables d'appraisal
Elaboration du processus d'appraisal	Faible	Faible	Très élevée
Phénomène adaptatif ?	Oui	Non	Oui
Consensus	Les émotions sont multi-composantes Les émotions ont un caractère inter-culturel Les émotions sous-tendent l'action Les émotions agissent comme organisateur de la prise de décision		

TABLE 1.1 – Tableau récapitulatif des modèles théoriques dominants de l'émotion.

Concepts-clé de la théorie des émotions

- Activation (arousal) : dimension fondamentale de l'expérience émotionnelle, avec la valence, dans le modèle bi-dimensionnel.
- Valence : expression de la dimension de plaisir/déplaisir de l'émotion.
- Cognition : Coppin et Sander en proposent une définition générique comme « un processus, naturel ou artificiel, qui traite [...] de l'information, servant à l'acquisition, l'organisation et l'utilisation de connaissances de manière explicite ou implicite[Coppin and Sander, 2010]. »
- Évaluation cognitive (appraisal) : évaluation cognitive, rapide et inconsciente, de stimuli ou événements.
- Tendance à l'action : Selon Frijda, [Frijda, 1986], état de préparation de l'individu « dans le but exécuter un certain type d'action. ».
- Émotion discrète (émotion de base) : selon la théorie évolutionniste, les émotions de base sont au nombre de 6 (tristesse, colère, dégoût, peur, intérêt, joie). Elles sont innées, universelles et invariantes tout au long de l'existence.

Chapitre 2

Analyse de l'expression faciale

2.1 Introduction

Après avoir posé les bases de la théorie des émotions dans le chapitre 2, nous abordons l'analyse automatique de l'expression faciale.

On peut relever quatre types de signaux parmi les informations relayées par le visage [Pantic and Bartlett, 2007] :

- *les signaux faciaux statiques*, permanents au cours de la vie, tels que la structure osseuse, les proportions, la forme du visage.
- *les signaux lents*, tels que les rides et ridules, qui se forment sur le long terme, ayant comme effet d'atténuer les zones de démarcation des éléments du visage.
- *les signaux artificiels*, tels que les lunettes, le maquillage ou les cosmétiques.
- *les signaux rapides*, issus de l'activité neuromusculaire du visage. Ils ont une temporalité et une intensité qui sont de nature à influer sur la signification qui leur est donnée de l'extérieur. Ils sont considérés comme signaux de mesure atomique de l'expression faciale au sens large.

Depuis les travaux de Paul Ekman [Ekman and Friesen, 1975], on sait que ce sont les signaux rapides qui témoignent de l'émotion. Ce sont ceux auxquels nous nous intéressons dans ce chapitre.

Comme la technique de reconnaissance de l'émotion est intimement liée à la théorie et à la méthodologie proposée par Paul Ekman, nous introduisons ce chapitre par une biographie, permettant de situer ses recherches.

Biographie de Paul Ekman

Paul Ekman

Paul Ekman est professeur de psychologie clinique de l'université Adelphi. Ses premières recherches portent sur le comportement non verbal à travers les cultures. C'est à cette occasion que lui viendra l'idée de tester sur des individus de différentes cultures la reconnaissance des émotions sur un visage. Voyageant d'abord au Japon, Brésil, Argentine avec ses photos, il propose l'universalité de l'identification à une émotion après avoir validé ses hypothèses sur des tribus éloignées de la Papouasie-Nouvelle Guinée.Il commence alors l'inventorisation, sur base de la musculature du visage, de toutes les expressions qu'un visage peut afficher. Sa rencontre avec Silvan Tomkins, professeur de psychologie à Princeton, le convainc que l'émotion est un processus clé et qu'il est possible de le déchiffrer. La définition du FACS prendra 7 ans et s'est depuis imposée comme standard de facto de la mesure du comportement facial. Poursuivant ses recherches sur les mouvements du visage, il découvre les micro-expressions, signaux fugaces et involontaires passant sur le visage, les micro-expressions permettraient de distinguer avec certitude le caractère spontané ou non d'une expression. Avec Mark Frank de la Rutgers University, il met sa technique à disposition du DARPA pour la détection du mensonge dans la lutte anti-terroriste. Paul Ekman est fondateur de plusieurs sociétés offrant des services de formation à la lecture faciale ou à la certification FACS. Il est également conseiller pour Emotient, startup rachetée en 2016 par Apple pour sa technologie de reconnaissance faciale de l'émotion.

L'objet de ce chapitre est d'opérer la distinction entre reconnaissance de l'expression faciale et reconnaissance de l'émotion.

La reconnaissance de l'expression faciale est un processus descriptif des mouvements et de la déformation des traits en unités abstraites d'action faciales.

Par contraste, la reconnaissance de l'émotion est un processus interprétatif d'un ou plusieurs facteurs, qui peuvent être observables ou non, à travers plusieurs canaux de communication tels que les mouvements des traits, le regard, mais aussi la voix ou le comportement du corps.

On ne peut donc pas concevoir de processus de reconnaissance de l'émotion si l'on ne prend pas en compte le fait qu'on ne peut l'observer directement, mais bien l'inférer à partir de l'expression, des données de self-report, du contexte ou des données physiologiques (voir Fig.2.1.1

).

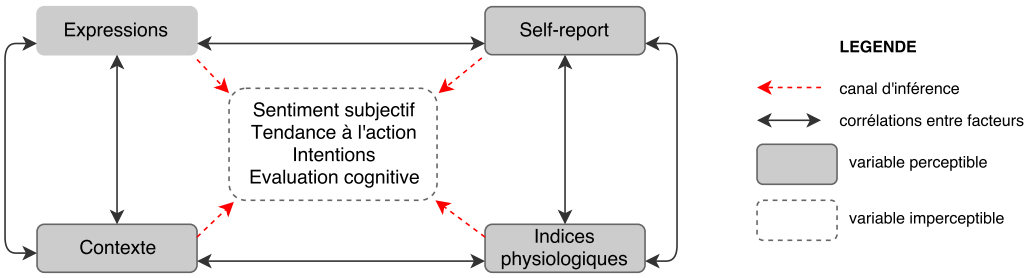


FIGURE 2.1.1 – Relation entre variables perceptibles et imperceptibles et inférence de l'émotion.

Source : adapté de [Cohn, 2006].

Dans ce chapitre, nous abordons l'expression faciale, qui ne constitue donc qu'un canal parmi d'autres de la mesure de l'affect, même s'il en est un des principaux.

En effet, des travaux d'Albert Mehrabian [Mehrabian, 2008], il ressort que les différents canaux (expression faciale, vocale, comportement, signaux physiologiques) n'influent pas de manière égale sur la transmission des signaux que nous envoyons à propos de nos émotions.

Les expressions faciale et vocale jouent un rôle prépondérant par rapport au contenu verbal et à la gestuelle. L'expression faciale compterait pour 55%, l'intonation pour 38%, pour 7% seulement pour le contenu.

Les signaux physiologiques étant peu pris en compte en raison de la difficulté à les capter en permanence.

Le chapitre est articulé comme suit : nous décrivons d'abord les deux approches conceptuelles qui prévalent en reconnaissance de l'expression faciale : la méthode par jugement et la méthode par la mesure des signes. Nous chercherons à en identifier les fondement et à comprendre comment elles sont traduites en systèmes fonctionnels.

Nous nous intéressons ensuite à la manière dont on peut décomposer la dynamique d'une expression faciale, la codifier en unités de mouvement minimales et en mesurer les variables de timing et d'intensité.

Enfin, nous illustrons ces deux approches en décrivant leur mise en œuvre, et nous comparons deux types de systèmes opérant à des niveaux différents : en reconnaissance de l'émotion et en détection de l'activité musculaire faciale.

Une des questions largement débattues en recherche sur la reconnaissance de l'émotion est de savoir s'il est plus efficace de procéder par détection directe des mouvements du visage, ou au contraire d'extraire d'abord, et de mesurer ces actions atomiques de mouvement, pour en déduire ensuite un label émotionnel.

Enfin, nous passons en revue les facteurs importants de la reconnaissance des émotions, dont la détection est innée chez l'être humain mais encore hors de portée des systèmes informatiques.

L'intensité de l'expression, son caractère posé ou spontané, ou les éléments de contexte, sont des éléments essentiels dans la démarche de doter un système d'intelligence artificielle d'une véritable sensibilité à l'émotion.

2.2 Mesure et interprétation de l'expression faciale

La recherche sur la détection des émotions peut être caractérisée par deux courants [Pantic and Bartlett, 2007] : la reconnaissance faciale de l'affect et la détection de l'activité musculaire faciale.

Ces deux méthodes prennent leur source dans deux approches fondamentales de mesure de l'expression en psychologie : l'approche par jugement (*message judgment*) et l'approche par signes (*sign-measurement*), inspirant respectivement reconnaissance de l'émotion et détection de l'activité musculaire faciale.

La reconnaissance faciale de l'affect a pour objet de classifier les expressions du visage en labels émotionnels (généralement limités aux 6 émotions de base : colère, tristesse,...), tandis que la détection de l'activité musculaire va chercher à décomposer une expression du visage pour en extraire les unités minimales, capturant éventuellement leur intensité et leur dynamique temporelle.

Ces unités, toujours découplées de toute interprétation subjective, sont appelées Action Unit (AU) ou des Facial Animation Parameters (FAP), selon que l'on utilise la codification Facial Action Coding System (FACS) ou FAP.

2.2.1 Base conceptuelle

Paul Ekman et Wallace Friesen ont conceptualisé la différence entre ces deux approches de mesure de l'activité faciale, qui, bien qu'elles portent sur des dimensions différentes et observent des phénomènes différents, peuvent parfois répondre aux mêmes questions.

Pour expliquer la différence entre ces deux approches, nous posons l'exemple d'un groupe de spécialistes, auxquels serait présenté un échantillon d'enregistrements des mouvements faciaux de patients diagnostiqués dépressifs. La question posée serait de savoir si l'état mental dépressif est bien lié à des patterns comportementaux du visage.

Selon l'approche par jugement, il serait demandé aux spécialistes d'inférer à partir du visionnage des images brutes, si oui ou non le patient est atteint de dépression.

Si le jugement des spécialistes amène à un diagnostic correct, il en est déduit que l'expression faciale témoigne bien de l'état mental dépressif du patient.

En utilisant la méthode par signes en revanche, c'est en examinant l'analyse du découpage précis des mouvements du visage, par exemple entre patients normaux et dépressifs, et en y

identifiant par la suite correctement les patterns dont on sait qu'ils dissocient les deux groupes, que l'on peut conclure que l'activité faciale est un marqueur d'un état dépressif.

La méthode par jugement se fonde sur la capacité à **inférer**, (à partir d'une expression) une émotion, des traits de personnalité ou des attitudes. Les observateurs sont nommés *juges*.

L'évaluation par jugement est holistique. Un (panel de) juge(s) va attribuer un label à chaque image, sur base de l'expression globale du visage. Ce label sera généralement une des 6 émotions de base.

La fiabilité de la base expérimentale se fondera sur le taux d'accord entre les réponses d'un échantillon de juges (experts et non-experts). Les labels généralement produits selon l'approche par jugement en reconnaissance faciale sont les 6 émotions de base décrites par Ekman.

Suivant la méthode par signes, l'observateur **décrit** le comportement facial, en termes d'occurrences des mouvements, de leur durée, de leur intensité, et des muscles impliqués, et on associe à ces unités de mouvement des classes abstraites.

La codification la plus couramment utilisée en méthode par signes est la codification FACS. Les observateurs sont plutôt appelés *codeurs* dans cette approche.

La nuance entre ces deux approches n'est pas seulement conceptuelle. La question est posée de savoir s'il est plus pertinent de détecter l'émotion en appliquant directement, sur une séquence d'images, un classifieur (entraîné avec un échantillon d'images labélisées), ou s'il est au contraire plus pertinent de passer d'abord par une première phase de détection des AUs, dont on déduira un label émotionnel, par exemple via un algorithme « rule-based », la correspondance avec une émotion.

La grande force de la théorie des émotions universelles proposée par Ekman, est justement de proposer une liste finie d'émotions, auxquelles sont liés des patterns d'activité faciale codés génétiquement. Il serait donc toujours possible d'associer toute combinaison de mouvements musculaires faciaux à un label émotionnel.

Pour les concepteurs de systèmes automatiques d'analyse de l'expression, ces deux méthodes revêtent une importance particulière, car elles influent sur deux paramètres importants de la reconnaissance de l'expression :

1) la capacité à disposer rapidement d'une base de connaissance large, variée et fiable pour entraîner les algorithmes,

2) la complexité du problème à traiter par l'algorithme.

Nous revenons sur le premier point à la section 3, où nous abordons la problématique de la reconnaissance automatique des AUs, et sur le second à la section 2.3.1, où nous passons en revue les deux modèles (détection en une phase de l'émotion Vs. détection en deux phases : des AUs puis de l'émotion).

De cette différence conceptuelle entre deux approches découlent deux conceptions différentes des systèmes de reconnaissance automatique de l'expression faciale : la *reconnaissance faciale de l'affect* pour la méthode par jugement, l'*analyse de l'activité musculaire faciale* pour la méthode par signes. Nous présentons une description haut niveau des deux approches ci-dessous.

2.2.2 Reconnaissance faciale de l'affect

La plupart des systèmes automatiques d'interprétation de l'affect se limitent à la détection des 6 émotions de base, hormis quelques travaux dirigés vers la reconnaissance d'états non émotionnels, tels que la fatigue, l'attention ou la douleur [Valstar et al., 2006].

Les taux de reconnaissance atteints sur base du corpus existant pour l'entraînement des systèmes sont élevés. Mais l'applicabilité aux conditions de la vie réelle est encore limitée pour les raisons suivantes :

1. Les émotions de base (ou prototypiques) seules ne couvrent qu'une fraction limitée des émotions que nous développons dans la vie réelle.
2. Dans la réalité de tous les jours, l'occurrence des émotions dites prototypiques est rare. Les émotions mixtes, subtiles ou mélangées sont beaucoup plus fréquentes.
3. L'apprentissage d'un système générant des labels émotionnels requiert une labélisation des images du dataset par l'approche par jugement, induisant un biais lié à la subjectivité du codeur.
4. Si l'universalité des émotions de base fait (jusqu'à un certain point) consensus, l'influence de l'environnement culturel et social est avéré pour les autres, ce qui rend les classes d'émotions difficilement transportables, et surtout difficilement généralisables à de nouveaux visages.

Une application de l'approche par jugement serait le recours aux techniques de machine learning pour entraîner des classifieurs sur une base de données d'images pourvues de labels marquant l'émotion élicitée.

C'est l'approche décrite par Sebe et.al [Sebe et al., 2007], avec l'avantage de la mise au point d'un scénario pour collecter une base de données d'images spontanées. Si le taux reconnaissance varie de 85,6% à 95,6%, le nombre d'émotions mesurées est limité à 6 (joie, surprise, colère, dégoût, peur et tristesse) et le processus de détection est difficilement généralisable, en raison de l'échantillon limité de sujets inclus dans la base de données d'entraînement.

On comprend bien par ailleurs que la validité de ce travail porte plus spécifiquement sur l'aspect technique, par la comparaison très fouillée de plusieurs techniques de machine learning telles que les Neural Bayesian Network (NBN), les SVM ou les arbres de décision.

2.2.3 Analyse de l'activité musculaire faciale

Dans la vie de tous les jours, la survenance des émotions prototypiques telles que les décrivent les tenants des émotions primaires, est rare.

L'émotion est le plus souvent affichée par des signes de faible portée, tel le froncement d'un sourcil témoignant d'un ressenti négatif.

Dans ces cas, disposer d'un système n'opérant qu'à plus haut niveau n'est pas utile. Seule les FACS peuvent rendre compte d'événements à ce niveau de finesse d'analyse [Cohn et al., 1999].

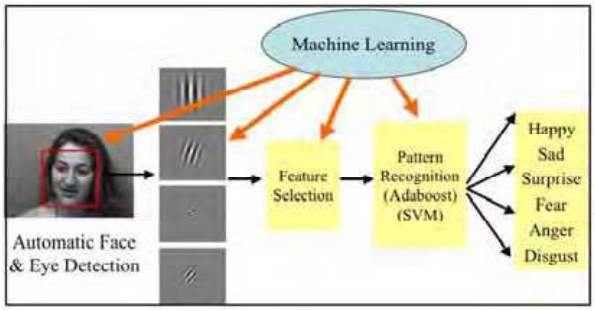


FIGURE 2.2.1 – Processus de détection par machine learning de label émotionnels

Source : [Pantic and Bartlett, 2007]

En parallèle à la détection de l'affect, une autre voie privilégie l'analyse de l'expression faciale par la détection de l'activité musculaire. L'analyse est portée à un niveau plus fin, les unités de mesure les plus courantes étant les FACS ou les FAP.

La détection des AUs plutôt que des émotions offre les avantages suivants :

1. Les Action Units, dans le cas de FACS, sont dé-corrélées de toute interprétation subjective, ce qui les rend utilisables pour tout type de mécanisme d'inférence de plus haut niveau à partir des signaux rapides (tels que décrits dans la taxonomie proposée par [Ekman and Friesen, 1969] :
   - Signaux affectifs / attitudinals : joie, peur, intérêt, stress,...
   - Emblèmes : signaux issus du contexte culturel,
   - Régulateurs : signaux médiateurs de la conversation, tels que le sourire,
   - Illustrateurs : signaux accompagnant la parole, tels que le mouvement du sourcil.
2. Les AUs sont particulièrement utiles comme paramètres de moyen-niveau car il permettent d'une part de réduire la totalité des configurations faciales différentes exprimables par la musculature du visage à 44 unités d'action. Ce sont aussi ces AUs qui permettent de faire un lien direct avec une émotion. La table de correspondance entre les 6 émotions universelles et leur(s) configuration(s) musculaires se trouve codifié dans le dictionnaire Emotional facial action coding system (EMFACS).
3. Le système FACS est compréhensible et accessible à travers les différents domaines. Il rend possible la détection de nouveaux patterns expressifs liés à des états mentaux particuliers. Par exemple quels sont les patterns d'expression faciale liés à la fatigue ? ou à la perte d'attention ? C'est par l'objectivation sur les traits du visage de ces marques de fatigue, que l'on peut ensuite espérer construire des systèmes dont la fiabilité sera également objectivable.
4. Enfin, il est un certain nombre de cas où la mesure par les FACS s'avère plus fiable que le jugement subjectif. Par exemple, pour distinguer la déception du dégoût, ou pour détecter le mensonge [Frank and Ekman, 2004].

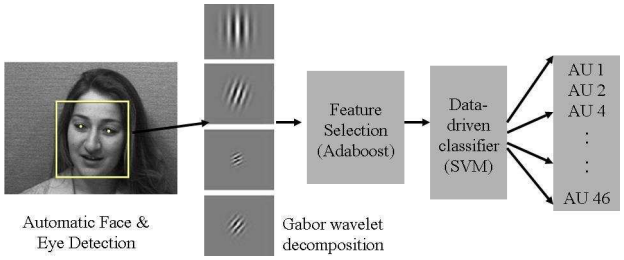


FIGURE 2.2.2 – Processus de reconnaissance des Action Units.

Source : [Pantic and Bartlett, 2007]

En résumé, FACS s'est imposé comme un corpus de connaissance largement partagé, pouvant servir à différentes applications, capitalisant sur une somme d'expérience accumulée sur le long terme.

Pour autant, l'approche par signes présente également ses faiblesses.

D'une part, de la même manière que les 6 émotions de base ne captent pas l'entièreté des émotions possibles, les systèmes de détection d'AUs actuels ne permettent encore d'en détecter qu'une trentaine sur les 44 existantes.

Ce sont celles qui sont le plus fréquemment observables, et pour lesquelles on dispose d'un échantillon de taille significative pour constituer des bases de donnés d'entraînement.

Ensuite, même s'il laisse moins de place à l'interprétation, le codage manuel des AUs à partir d'une séquence vidéo reste un travail fastidieux et coûteux.

L'automatisation du codage FACS par les techniques de machine learning est un thème important de recherche.

Fast-FACS par exemple, [De la Torre and Cohn, 2011] propose un système d'aide au codage où l'intervention d'un codeur se limite à coder l'apex de l'expression et son intensité, réduisant le temps de codage de 50%.

Tian *et.al* [Tian et al., 2005] résument ainsi les 4 axes principaux de développements récents dans ce domaine :

a) l'augmentation de la multiplicité des traits (*features*) qu'il est possible d'extraire en temps réel,

b) un focus sur la détection d'AUs (ou combinaisons d'AUs), plutôt que sur les labels émotionnels plus complexes,

c) un meilleur rendement des techniques d'acquisition, d'extraction, et de représentation des *features* par les techniques de machine-learning,

d) le développement de systèmes totalement automatiques (capables d'opérer, après apprentissage, sans l'intervention d'un opérateur) de détection en temps réel.

Pour bien comprendre que les deux approches ne sont pas hermétiques, dans la section 3.4 ci-

après, nous abordons une question importante et largement traitée : est-il plus efficace d'extraire les émotions directement des images à bas niveau, grâce à des classifieurs entraînés sur des bases d'images labellisées, ou bien faut-il plutôt effectuer une détection en deux passes, en passant par une phase intermédiaire de détection des AUs, pour en inférer ensuite de l'émotion ?

2.2.4 Codification de l'expression faciale

Jusqu'à la définition du FACS en 1977, les recherches comportementales sur l'expression faciale étaient basées sur l'observation des traits de manière holistique, par des juges, avec leur part de subjectivité.

Leur évaluation pouvait dépendre du contexte, des niveaux de précédence donnés à l'expression faciale, vocale, au comportement, ou de leur contexte socio-culturel.

L'idée d'objectiver les actions du visage en un corpus standardisé remonte aux années 1920, mais deux standards se sont imposés depuis : le système FACS proposé par Ekman et Friesen, et les FAP, qui font partie du standard MPEG-4 Synthetic/Natural Hybrid Coding (SNHC).

L'apport de cette méthodologie est de découper l'observation d'une action du visage en éléments atomiques de mesure de la déformation du visage. Le visage n'est plus évalué de manière holistique, mais chacun des mouvements musculaires est localisé, et codé dans une nomenclature adhoc.

La codification faciale répond également à un besoin de disposer d'un fondement objectif permettant aux différents systèmes de se comparer en utilisant une représentation commune.

Une des clés de la reconnaissance de l'émotion est la codification, pour chaque expression, de ses configurations temporelles et quantitatives. Or, ceci a été fait manuellement jusqu'ici, par des experts codeurs, en visionnant et marquant les extraits vidéo, image par image.

Un des thèmes de recherche les plus actifs porte sur l'automatisation de la détection des mouvements dans les séquences vidéo.

Être capable d'extraire les unités d'action d'une séquence, avec leur intensité et leur développement temporel, constitue un pas en avant important dans l'étude des sciences comportementales.

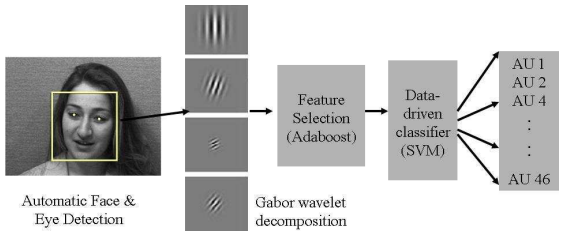
Ceci aurait pour effet, en plus de rendre accessible la mesure objective du comportement facial, de stimuler la découverte de nouveaux patterns de comportement faciaux, et ce avec des techniques permettant de surpasser le jugement humain en rapidité et en précision.

Dans le cadre de la reconnaissance de l'émotion, il existe une tendance croissante des groupes de recherche à aller vers la détection de l'activité musculaire plutôt que vers l'identification de labels d'expressions prototypiques, comme les 6 émotions de base.

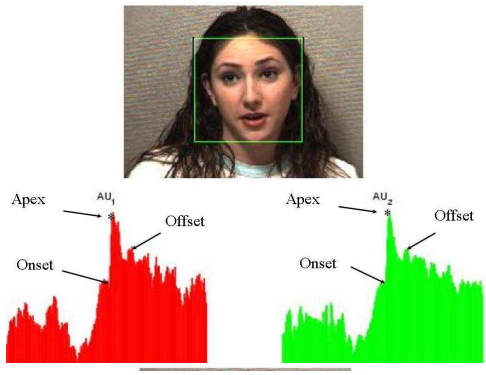
Il y a une volonté de faire abstraction de la dimension subjective du choix des labels, au profit de critères plus objectifs.

L'implémentation d'un tel système est donnée en exemple dans [Bartlett et al., 2006]. Le système détecte de manière totalement automatique 26 unités de mouvement sur 44, fournissant

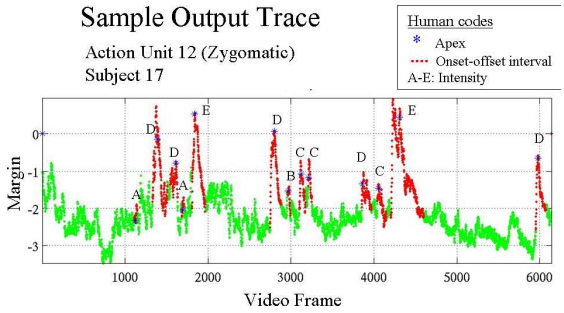
les marqueurs temporels et l'intensité (Fig.2.2.3).



(a) Architecture générale du système.



(b) Exemple d'output pour une séquence de 10 secondes des AUs 1+2 (relèvement de sourcil).



(c) Exemple d'output pour l'AU 12 (sourire). Le résultat inclut l'intensité ainsi que les onset-apex-offset.

FIGURE 2.2.3 – Le système de reconnaissance automatique des AUs proposé en [Bartlett et al., 2006].

2.2.5 Facial Action Coding System (FACS)

Le FACS associe chaque déformation du visage, par les muscles qui l'activent ou chaque déformation de la peau, à des « unités de mouvement », les Facial Action Unit (FAU). Il existe 44 AUs, qui sont les plus petites unités d'activité faciale discernables. Le FACS produit également les règles qui prévalent au codage des mouvements du visage par des observateurs humains, ainsi que leur dynamique temporelle. Pour la liste des 44 AUs, il est possible de se référer à [Cohn et al., 2007].

A partir des 44 Action Units, des combinaisons sont formées pour décrire des expressions plus

AU 1	AU 2	AU 4
		
Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together
AU 5	AU 6	AU 7
		
Upper eyelids are raised.	Cheeks are raised.	Lower eyelids are raised.
AU 1+4	AU 4+5	AU 1+2
		
Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.	Inner and outer portions of the brows are raised.
AU 1+2+4	AU1+2+5+6+7	AU0(neutral)
		
Brows are pulled together and upward.	Brow, eyelids, and cheek are raised.	Eyes, brow, and cheek are relaxed.

FIGURE 2.2.4 – Action Units de la partie supérieure du visage

Source : Tian, Y.-L., Kanade, T., & Cohn, J. F. (2001). Recognizing Action Units for Facial Expression Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), 97-115.

évoluées. Ekman évalue le nombre de d'expressions faciales différentes pouvant être affichées sur un visage humain par combinaison de ces 44 AUs, à 7000. Il existerait donc 7000 « configurations faciales »(représentant un éventail plus large que les 6 émotions de base, une seule émotion pouvant également revêtir de nombreuses variantes).

Le FACS est indépendant de toute notion d'interprétation. Il est purement descriptif et ne porte aucun label émotionnel. Le lien entre les AUs et leur interprétation émotionnelle est extrinsèque au standard.

La base empirique de cette correspondance est décrite dans des systèmes de codification externes, tels qu'EMFACS ou FACSAID, où se retrouvent la plupart des expressions prototypiques et leurs combinaisons en termes d'AUs.

Il est à noter que l'utilisation des FACS ou FAP n'est pas exclusive aux techniques de reconnaissance automatique.

Leur palette d'utilisation est plus large : en sciences comportementales, pour le traitement de la schizophrénie ou de la dépression, mais aussi en zoologie, en éthologie, en science de l'éducation ou dans le domaine de l'animation graphique.

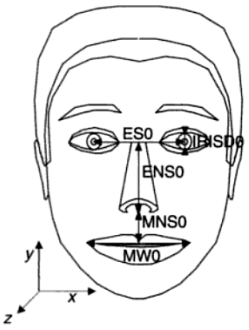
Le principal apport des FACS est de proposer une codification compréhensible, qui établit un lien entre des configurations physiques du visage d'une part, et ce que l'on en découvre par l'étude des sciences comportementales d'autre part.

La découverte de nouveaux patterns d'expression est rendue possible par leur description en éléments de mesure objectifs.

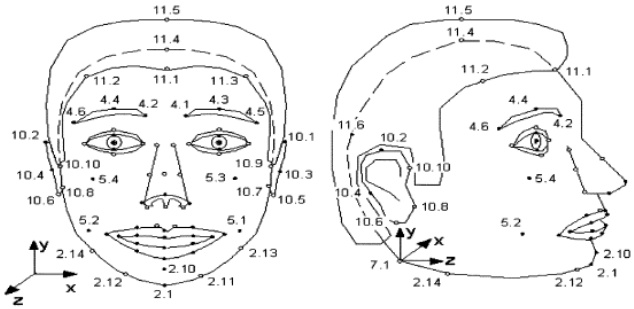
Par exemple, objectiver la mesure du niveau d'attention chez un opérateur machine ou un conducteur, ou qualifier les signes d'attention ou de distraction pour un système de tutoring intelligent.

Le FACS repose sur la capitalisation du travail fait en amont par l'ensemble de la communauté des sciences comportementales.

Le principal inconvénient est le processus de codage manuel sur base de séquences vidéo analysées image par image, qui est coûteux en temps et en ressources.



(a) Mesure des FAP Units (FAPU) sur un visage neutre.



(b) Positionnement des Feature Points (FP) sur le visage neutre.

FAP No.	FAP Name	FAP Description
3	open_jaw	Vertical jaw displacement
4	lower_t_midlip	Vertical top middle inner lip displacement
5	raise_b_midlip	Vertical bottom middle inner lip displacement
6	stretch_l_cornerlip	Horizontal displacement of left inner lip corner
7	stretch_r_cornerlip	Horizontal displacement of right inner lip corner

(c) Extrait de la liste des 84 FAPs

FIGURE 2.2.5 – Métriques des Facial Animation Parameters (FAP)

2.2.6 Facial Animation Parameters (FAPs)

Le développement des FAP procède de la même logique que les FACS, hormis qu'ils tirent leur origine de la recherche sur les techniques d'animation plutôt que sur les études comportementales. C'est le groupe Moving Pictures Experts Group (MPEG) qui est à l'origine du standard MPEG-4, dont les FAP font partie.

La codification FAP spécifie un visage neutre (dont les muscles sont relâchés, paupière ouverte,...) et 84 Feature Point (FP), qui représentent chacun un point du visage pouvant faire l'objet d'une déformation, liée à une métrique, le FAPU (fig.2.2.5b).

Cette intensité est exprimée sous forme de fraction de la distance entre les FP (fig.2.2.5a), ce qui permet d'en caractériser l'intensité.

84 FAP (fig.2.2.5) permettent donc, de manière similaire aux 44 AU, de décrire (ou de spécifier), par combinaison, toute déformation sur le visage, référence gardée par rapport au visage neutre.

Si historiquement, le standard MPEG-4 est plus orienté vers l'animation et la synthèse d'expressions faciale, les deux systèmes sont liés : il est possible d'établir la correspondance entre les deux schémas. Un système qui permet, par le mapping entre FACS et FAP, d'allier deux modules de reconnaissance d'émotion et de synthèse de l'émotion est proposé en [Tsapatsoulis et al., 2002].

Ce travail est basé à la fois sur les recherches empiriques en psychologie et sur l'expérimentation par les bases de données, notamment Medialab (la base de données du MIT Affective Computing group).

Les applications d'une telle solution sont tournées soit vers l'Interaction Homme-Machine (IHM), avec la possibilité de s'adapter dynamiquement à l'état émotionnel de l'humain interagissant avec lui, soit vers la synthèse d'agents intelligents, capables d'enrichir le contenu du message par des rendus affectifs.

2.3 Schéma d'un système d'analyse de l'expression faciale

Les bases de la construction d'un système de détection de l'émotion peuvent dès lors se résumer en 3 modules principaux (Fig. 2.3.1) :

- Acquisition du visage : cette phase consiste à la détection et au redimensionnement de l'image, de sorte à isoler un visage du paysage, et de tout ce qui pourrait être considéré comme du bruit à l'analyse.
- Extraction des traits : cette phase a pour but d'isoler la représentation des traits significatifs du visage (les yeux, la bouche, les sourcils, la texture de peau,...)
- La classification : c'est la phase au cours de laquelle un algorithme (typiquement un classifieur entraîné sur une base expérimentale) va étiqueter une image ou une séquence avec un label, qui peut être un mouvement musculaire, une émotion prototypique, ou une comportement attitudinal (douleur, fatigue,...).

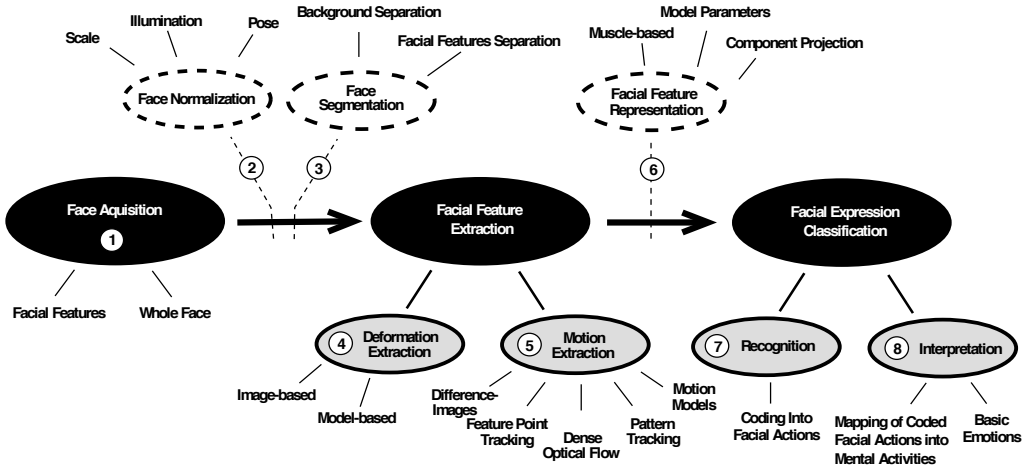


FIGURE 2.3.1 – Un système générique de reconnaissance de l'expression faciale.

Source : Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis : a survey. Pattern recognition, 36(1), 259-275.

Dans notre travail, nous n'abordons pas les 2 premières phases, qui font appel à un large éventail de techniques propres au traitement de l'image qui ne sont pas liées directement à la reconnaissance des émotions.

Pour donner corps à la différence de concept entre méthode par jugement et par signes, nous illustrons chacune des deux méthodes par un exemple ci-dessous, ensuite nous évaluons l'hypothèse que la détection des AUs en préalable à la reconnaissance de l'émotion peut améliorer l'efficacité d'un système.

2.3.1 Détection directe d'émotions de base

Nous illustrons ici la méthode par jugement (Fig.2.3.2).

Sebe *et al* [Sebe et al., 2007] proposent un système de reconnaissance directe des 6 émotions de base. Le système est entraîné sur une base de données dont chacune des séquences porte le label d'une seule émotion, le taux de reconnaissance sur les 6 émotions élicitées va de 86% à 96%. La base de données a été constituée à partir d'images spontanées prises à partir de contenus vidéo.

La vérité de terrain est ici constituée par le codage des labels d'émotion (non par des codeurs mais par l'interrogation des participants sur l'émotion ressentie).

A partir de chaque image, un vecteur de mouvement, Motion Unit (MU) est généré. Le MU représente la direction et l'intensité du mouvement de chaque trait. Ce vecteur de mouvement sert d'input à un classifieur, qui aura au préalable été entraîné sur la base contenant les images munies des labels émotionnels.

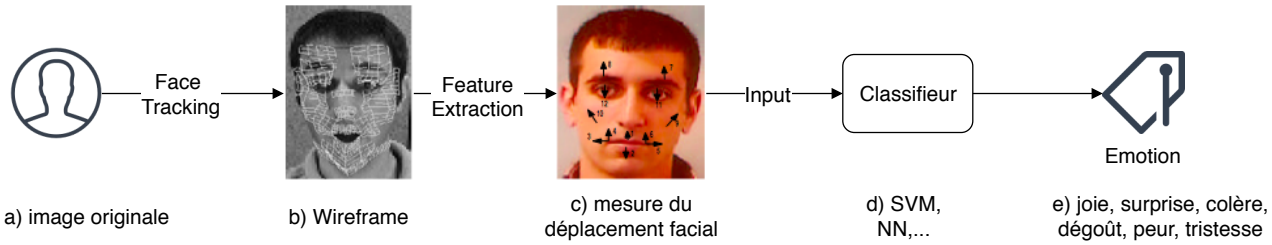


FIGURE 2.3.2 – Reconnaissance directe d'émotion par Machine Learning [Sebe et al., 2007].

Dans ce cas de figure, un classifieur opère sur une base expérimentale qui est validée par le jugement des intervenants, l'accent est plutôt mis sur le type de classifieur à utiliser pour inférer au mieux du label émotionnel à donner à chaque image de la base de test.

2.3.2 Identification d'une émotion par détection des Action Unit

Dans une approche par signes (Fig. 2.3.3), les AU's sont utilisés comme paramètre de moyen-niveau en input d'un système d'inférence de plus haut niveau.

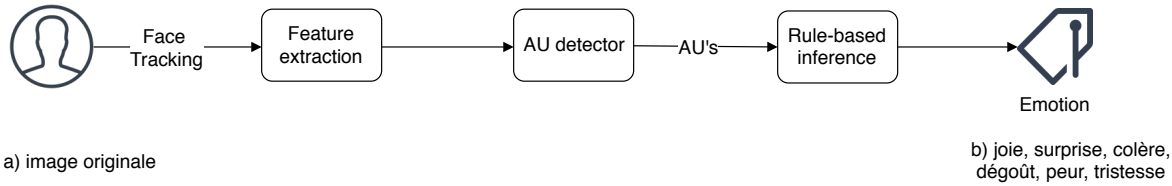


FIGURE 2.3.3 – Reconnaissance de l'émotion après détection des Action Units, par inférence sur une table de règles.

Ici en l'occurrence, l'inférence du label émotionnel se fait sur base d'un système expert muni d'une table faisant correspondre les Action Unit et les émotions (Fig. 2.3.4).

AUs	Emotion	AUs	Emotion	AUs	Emotion	AUs	Emotion
1 + 2	Surprise	1	Sadness	23 + 17	Anger	10 + 17	Disgust
2	Anger	4	Anger	23 + 26	Anger	10 + (25/26)	Disgust
6	Happiness	5	Surprise	23	Anger	10	Disgust
1 + 4 + 5 + 7	Fear	7	Anger	24 + 17 + 26	Anger	9 + (25/26)	Disgust
1 + 4 + 5	Fear			24 + 17	Anger	9 + 17	Disgust
1 + 4 + 7	Sadness	27	Surprise	24 + 26	Anger	9	Disgust
1 + 5 + 7	Fear	20 + (25/26)	Fear	24	Anger	12 + (25/26)	Happiness
1 + 4	Sadness	20	Fear	10 + 16 + (25/26)	Anger	12	Happiness
1 + 5	Fear	15 + (25/26)	Sadness	10 + 17 + (25/26)	Disgust	16 + (25/26)	Anger
1 + 7	Sadness	15	Sadness	9 + 17 + (25/26)	Disgust	17	Sadness
5 + 7	Fear	23 + 17 + 26	Anger	12 + 16 + (25/26)	Happiness	26	Surprise

FIGURE 2.3.4 – Table d'inférence du système ISFER.

Source : [Pantic and Rothkrantz, 2000]

D'autres systèmes utilisent d'autres métriques de représentation que les FACS. [Ioannou et al., 2005] utilise les FAP, et positionne l'émotion détectée sur un *circumplex* inspiré de la théorie de Russel (Fig. 2.3.5).

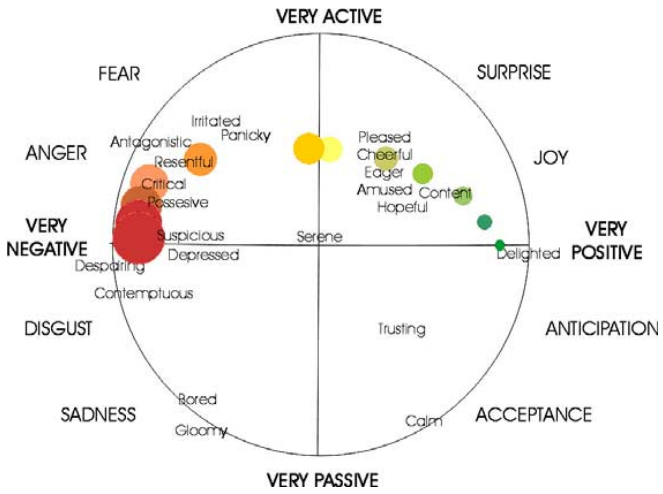


FIGURE 2.3.5 – Catégorisation des labels émotionnels sur le *circumplex*.

Source : [Ioannou et al., 2005]

L'inférence du label émotionnel se fait sur base d'un liste de règles liant FAP et position sur le cadran (Fig.2.3.6).

Rule	FAP	Quadrant
1	F3_H+F4_L+F5_VL+[F53+F54]_H+[F19+F21]_H+[F20+F22]_H	(+,+)
2	F3_M+F4_L+F5_L+[F53+F54]_H+[F19+F21]_H+[F20+F22]_H	(+,+)
3	F3_M+F4_L+F5_H+[F53+F54]_H+[F19+F21]_H+[F20+F22]_H	(+,+)
4	F3_H+F4_L+F5_L+[F53+F54]_H+[F19+F21]_H+[F20+F22]_H	(+,+)
5	F3_L+F4_M+F5_H+[F53+F54]_M+[F19+F21]_H+[F20+F22]_H+F31_M+F32_M+F33_M+F34_M+F37_M+F38_M+F59_H+F60_H	(+,+)
6	F3_H+F4_L+F5_VL+[F53+F54]_M+[F19+F21]_H+[F20+F22]_H	(+,+)
7	F3_L+F4_L+F5_H+[F53+F54]_H+[F19+F21]_H+[F20+F22]_H+[F37+F38]_M+F59_H+F60_H	(+,+)
8	F3_H+F5_VL+[F53+F54]_L+[F19+F21]_L+[F20+F22]_L+F31_H+F32_H+F33_H+F34_H+F35_H+F36_H+F37_L+F38_L+[F37+F38]_L	(+,+)
9	F3_H+F5_VL+[F53+F54]_M+[F19+F21]_L+[F20+F22]_L+F31_H+F32_H+F33_H+F34_H+F35_H+F36_H+F37_L+F38_L	(+,+)
10	F3_M+F5_L+[F53+F54]_L+[F19+F21]_L+[F20+F22]_L+F31_H+F32_H+F33_H+F34_H+F35_H+F36_H	(+,+)

FIGURE 2.3.6 – Table de correspondance (extrait) entre FAPU et position sur le *circumplex*.

Source : [Ioannou et al., 2005]

La Fig.2.3.6 illustre un moteur fournissant une liste de règles parcoures séquentiellement. Chaque règle contient une configuration de FAP et une position sur le cadran du circumplex.

Par exemple, la règle 1 fait correspondre une configuration de visage marquée par les FP 3, 4, 5, 19, 20, 21, 22, 53, 54 et leurs positions (High, Low, Very Low,...) à une position (+,+) sur le cadran, ce qui peut être assimilé à la joie ou la surprise.

2.3.3 Détection de l'émotion en deux phases

Les avantages à utiliser les AUs plutôt que de faire de la détection directe d'émotions en recourant aux techniques de machine learning sont les suivants ;

1. S'affranchir du caractère subjectif du jugement,
2. Faire abstraction du contexte socio-culturel qui pèse dans toute évaluation subjective,
3. Réduire la dimension du problème à résoudre à 44 unités,
4. Diminuer le temps et les ressources nécessaires à entraîner les modèles.

Par ailleurs, il existe un clivage en psychologie entre l'école classique et les études alternatives. La première postule, comme le fait EMFACS, qu'il existe une correspondance entre un nombre fini d'émotions de base et leur expression faciale. La seconde, inspirée des concepts de « Thin slices of behaviour »[Ambady and Rosenthal, 1992], qu'il existe un « quelque chose »dans la nature des individus qui permet à autrui de leur attribuer des attributs comportementaux ou affectifs.

Ce « quelque chose »d'expressif ne serait ni encodé ni décodé de manière volontaire ou consciente. Pourtant ils le sont avec une grande efficacité.

Nous n'approfondissons pas ici le champ des études modernes en psychologie, la référence étant donnée dans le seul but de rendre l'univers conceptuel de l'expérimentation.

Pour les concepteurs de systèmes informatiques, une des questions qui en découle est de savoir si ce processus peut être modélisé de façon plus efficace en utilisant une base finie de règles, conformément à la théorie classique, ou bien une structure de type réseau de neurones, plus proche de la seconde conception.

Valstar et.al [Valstar and Pantic, 2006] conduisent une double expérimentation destinée à tester si :

1. La détection des AUs préalablement à la détection de l'émotion en améliore l'efficacité,
2. Sur base des AUs détectées, une inférence de type « rule-based », comme le suggèrent les classiques, ou de type Neural Network est la plus à même de faire la transcription en un label émotionnel.

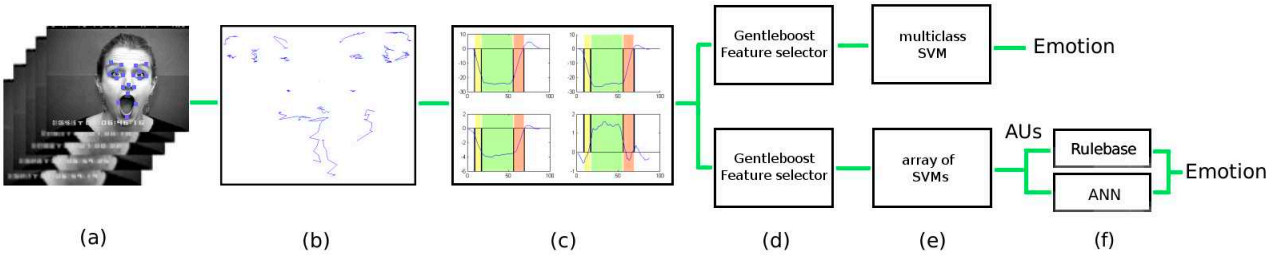


FIGURE 2.3.7 – Schéma d'expérimentation de la reconnaissance d'émotions par apprentissage direct et par détection des AUs au préalable. (a) image originale, (b) localisation des traits (c) extraction des 4 principaux traits impliqués par l'AU 2, avec leur timing (d) Sélection des traits principaux (e) injection dans un SVM (f) approche en deux phases : comparaison en classification par règles ou par Neural Network.

Source : [Valstar and Pantic, 2006]

Les traits du visage sont d'abord extraits, leur timing est inclus, et un algorithme appelé Gentleboost est utilisé pour sélectionner les traits les plus représentatifs à fournir au classifieur.

Deux configurations différentes sont ensuite expérimentées : dans un premier cas (approches en une phase), les traits sont classifiées par un SVM entraîné sur le database Cohn-Kanade.

Dans le second cas (approche en 2 phases), les traits sont d'abord injectés dans un SVM pour en identifier les AUs, ces AUs sont ensuite utilisés comme input d'une base de règles pour modéliser l'approche psychologique classique, d'un Neural Network (NN) pour adopter une représentation plus proche d'un processus inconscient.

Les résultats permettent de tirer les conclusion suivantes :

1. L'approche en deux phases est légèrement moins efficace en pourcentage de reconnaissance, mais ceci est compensé selon les auteurs par la complexité réduite du problème à 44 AUs et par l'abstraction du paramètre subjectivité.

2. La classification par base de règles se comporte significativement moins bien lorsque le détecteur d'AU enregistre des écarts par rapport à la codification manuelles. Ce qui s'explique par le fait que la méthode par règles est statique et n'apprend pas.

En conclusion, la méthode en deux phases est privilégiée, car le faible coût en précision de détection est compensé par une complexité sensiblement plus faible, et le fait qu'à plus grande échelle, il est impossible de collecter une base de donnée d'entraînement qui prenne en compte les 7000 expression possibles.

L'expérimentation validerait donc également la théorie alternative et l'approche en réseaux de neurones, par rapport à l'analyse par règles séquentielles.

2.4 Conclusion

Il existe deux grands courants de recherche en analyse de l'expression faciale : la détection de l'émotion et la détection de l'activité musculaire faciale.

Ces deux axes découlent des fondements issus de la recherche en psychologie, qui identifie deux approches pour la mesure du comportement facial : l'approche par jugement et l'approche par signes.

L'approche par jugement est interprétative : il s'agit, pour des juges, d'inférer la signification d'une expression faciale et de lui attribuer un label subjectif, généralement une émotion.

L'approche par signes en revanche est essentiellement descriptive, et agnostique à toute notion d'interprétation subjective.

Elle analyse une expression faciale en la décomposant en unités atomique de l'activité faciale. Image par image, ces unités de mouvement sont identifiées, et codées avec leur intensité et leur développement temporel.

L'approche par jugement se focalise très largement sur la détection des 6 émotions de base, tandis que l'approche par signes est dominée par la codification FACS, qui se présente comme une liste de 44 Action Units, ou plus petite unité de mouvement facial discernable, décomposant l'ensemble des configurations possibles chez l'humain en près de 7000 combinaisons.

Le standard FACS prévoit une méthodologie, un processus de certification et de codage du visage humain.

Si le taux d'accord inter-juge est élevé, il reste des inconnues sur la cohérence des observations entre laboratoires différents par exemple.

Les AUs ont leur utilité comme input dans des systèmes de décision de plus haut niveau, pour détecter des états plus complexes tels que les emblèmes, les régulateurs ou les illustrateurs.

A ce jour, on peut encore faire la critique aux systèmes existants qu'il existe une marge entre la détection sur base d'images posées et exagérées, et la vie réelle, où se mêlent des expressions mixtes, subtiles, et spontanées.

Un autre paramètre important est le coût important, en temps et en ressources, nécessaire pour coder manuellement les bases de données d'entraînement, tel que spécifié dans les FACS. L'extraction automatique des AUs est un axe de recherche important, particulièrement pour en capturer la dynamique temporelle, qui manque encore de maturité.

Il existe un corpus de connaissance croissant allant dans le sens que la dynamique temporelle est importante pour la détection d'états spécifiques comme la douleur ou l'humeur, mais aussi pour juger de la spontanéité d'une expression.

Entre approche par jugement et approche par signes, la seconde permet une réduction à 44 unités de la complexité du problème, et est prisée par la communauté informatique car dénuée de contexte culturel et subjectif.

Chapitre 3

Construction de la base expérimentale

3.0.1 Introduction

Un des points critiques dans l'élaboration de systèmes de reconnaissance automatique de l'émotion est de disposer de bases de données d'images labellisées, tant pour entraîner les modèles et les tester, que pour établir des bases de comparaison objectives entre différentes techniques.

L'existence d'un corpus de données étendu, disponible publiquement, et dont la validité fasse consensus, est un point qui fait défaut à ce jour dans la recherche sur la reconnaissance de l'émotion.

Il existe de nombreux champs d'application en recherche comportementale qui nécessitent la création de bases expérimentales.
Or, les équipes de recherches dans chaque discipline ont tendance à orienter la conception de leur corpus en fonction de leurs besoins spécifiques, avec comme conséquence une multiplication des sources de données.

Ainsi par exemple, la recherche en sciences informatiques dirigée vers la reconnaissance automatique de l'émotion tendra à développer des bases de données reprenant un grand nombre de sujets représentant un petit nombre d'émotions, tandis qu'en psychologie, où l'objectif sera plutôt de chercher des méthodes de différentiation plus fines entre les émotions, on privilégiera un plus petit nombre de sujets représentant un éventail d'émotions plus subtiles [Bänziger et al., 2012].

Une base de données doit répondre aux besoins suivants : 1) afficher une illustration de l'émotion la plus fidèle possible à celle que donnerait un visage dans la vie réelle, b) contenir pour chaque image, les méta-données permettant de l'associer à une Action Unit ou un label émotionnel, c) justifier d'un protocole permettant de valider la fiabilité du jugement ayant amené à l'attribution de ces labels, et d) prendre en compte les différences inter-individu et inter-cultures.

Dans ce chapitre, nous analysons d'abord les facteurs identifiés par les sciences comportementales comme influençant la signification d'une expression faciale, et dont doivent par conséquence tenir compte les concepteurs des bases de données.

Ensuite, nous détaillerons les principes sur lesquels sont construits les corpus de données existants. Nous détaillons également quels protocoles permettent d'en objectiver la fiabilité.

Enfin, nous évoquons la détection automatique d'Action Unit en remplacement du codage manuel.

3.1 Facteurs déterminants de l'analyse de l'expression faciale

3.1.1 Expression spontanée ou posée

La distinction entre expression faciale spontanée (réflexe non contrôlé face à un stimuli) et posée (jouée sous contrôle d'un instructeur) est essentielle pour un système de reconnaissance de l'affect.

La distinction trouve son origine dans le fait que ces deux types de comportement facial sont activés par deux régions distinctes du cerveau. D'autre part, les expressions commandées par ces deux régions diffèrent de par les muscles faciaux impliqués, et de par leur dynamique [Pantic and Bartlett, 2007].

Selon ce schéma, les groupe des expressions spontanées correspondrait à des expressions harmonieuses, cohérentes, apparentées aux réflexes. On y retrouve un petit groupe d'émotions de base.
A contrario, le groupe des expressions affichées de manière délibérée est caractérisé par des expressions moins souples, suivant des dynamiques temporelles plus variables, et serait apparenté aux expression issues de l'apprentissage culturel.

Les premiers datasets furent conçus sur base d'images fixes d'expressions posées. L'adaptation des systèmes aux conditions de la vie réelle imposent de développer la prise d'images spontanées dans les bases de données.

Sur l'ensemble des datasets existants, on peut identifier trois modes d'élicitation de l'émotion chez les sujets participants [Bänziger et al., 2012] :

- *Prototype posing* : les sujets sont amenés à afficher des expressions d'émotion déclinées exactement sous la forme de combinaisons d'AUs telles que spécifiées dans le modèle de Paul Ekman. Les acteurs sont soit guidés par un expert FACS, soit experts FACS eux-même. Le résultat est obtenu par répétition, jusqu'à former l'expression faciale conforme au standard FACS. Le ressenti réel de l'émotion n'est pas nécessaire et en conséquence, la validité viendra du fait que toutes la même émotion sera toujours représentée à l'identique par tous les acteurs dans toutes les images.

- *Communication effect acting* : dans ce scénario, les sujets sont aidés à exprimer une émotion sur base d'un mot, d'un concept ou d'un scénario dans lequel l'acteur puisera un contexte favorisant l'élicitation de cette émotion.
- *Felt experience enacting* : cette méthode privilégie l'authenticité de l'expression, en ayant recours soit à des techniques d'élicitation directe (par la projection de contenu vidéo par exemple), soit en invitant le sujet à revivre mentalement des scénarios dans lesquels une émotion particulière a été ressentie. Le résultat sera typiquement un éventail d'expressions plus variées entre individus que selon la première technique.

Le choix qui a le plus souvent été fait dans la constitution des bases de données existantes, est de privilégier les images posées, par *prototype posing*.

Ceci s'explique par le fait qu'en machine learning, un élément déterminant de l'efficacité d'un modèle est de disposer d'un grand nombre de sujets, et que la mise en place, le recrutement, la création de scénarios d'élicitation spontanée posent des contraintes trop lourdes en temps et en coût.

3.1.2 Dynamique temporelle de l'expression

L'analyse automatique de l'expression ne se limite pas à la détection instantanée d'une configuration musculaire ou géométrique. Un des grands points de convergence entre domaines de l'informatique et de la psychologie est de considérer l'émotion comme un phénomène évolutif et non instantané [Parkinson, 2005].

La mesure de la temporalité est précisément le facteur permettant une distinction plus fine des cas suivants :

- Expression spontanée ou expression posée.
- États complexes, comme la douleur ou l'ennui.
- Comportements socialement colorés, comme l'inhibition, la gêne ou la culpabilité.

Ainsi par exemple, c'est par la mesure de la dynamique que l'on peut faire la distinction entre une sourire « authentique »(appelé sourire de Duchenne, du nom du neurologiste éponyme, qui en avait déjà, au XIX$^{\text{ème}}$ siècle, identifié les patterns faciaux), et un sourire « social ». Le sourire de Duchenne est d'amplitude plus faible, possède des onset et offset plus courts, et sa durée totale est plus limitée.

3.1.3 Mesure de l'intensité

L'intensité d'une expression se mesure au degré de déformation par rapport au visage neutre. Dans le cas du sourire de Duchenne, ce serait le degré de relèvement des deux coins de la bouche par rapport à la situation neutre, matérialisant l'action du zygomatique majeur, codifié comme une action de mouvement minimale.

L'expérimentation montre qu'il existe une relation linéaire entre d'une part la précision à identifier un état mental et l'intensité avec laquelle il est perçu, et d'autre part l'intensité physique de l'expression du visage.

C'est ici encore un point important dans la distinction entre comportement spontané et posé.

La codification de l'activité faciale quelle qu'elle soit, doit donc prévoir une échelle à niveaux pour la mesure de l'intensité d'un mouvement.

Le standard FACS propose une échelle en 5 niveaux, de A à E qui quantifie une expression du point neutre à l'intensité maximale.

De nombreuses bases de données ont été constituées. [De la Torre and Cohn, 2011] en dresse un comparatif, en termes de nombre de sujets représentés, de scénario d'élicitation de l'expression, de type d'images et de prise de vues, et labels fournis dans les méta-données (des labels d'émotion dans une approche basée sur le jugement, ou des AU dans l'approche par signes).

Parmi les datasets présentés à la fig.3.1.1, la base de données Cohn-Kanade et sa version améliorée, Cohn-Kanade+, sont de loin les plus utilisées, avec toutefois les faiblesses suivantes :

- La prise d'image s'arrête à l'apex de l'expression.
- Ne contient aucune balise temporelle de l'expression.
- Ne combine pas images statiques et séquence vidéo.

3.2 Principes de construction d'une base de données idéale

Le but d'une base de données idéale serait de fournir une base d'apprentissage la plus large possible à un algorithme de classification.

La précision d'un classifieur augmente avec le nombre de données sur lequel on l'entraîne. Mais on veut également que le modèle que l'on va entraîner prenne en compte la plus grande diversité des expressions émotionnelles qu'on trouverait dans le monde réel.

Un même label d'émotion peut être assigné à des expressions faciales sur des sujets différents, dans des conditions de prise de vue différentes, sur des durées différentes, etc.

Mesurer l'efficacité d'un classifieur en termes absolus ne nous dit pas grand'chose, si l'on ne peut pas juger de sa capacité de généralisation.

Nous passons en revue les principes qui sous-tendraient la construction d'une base de données idéale.

3.2.1 Les critères techniques

Il y a un avantage à combiner images statiques et dynamiques. Les images statiques permettent d'assimiler avec précision la configuration de chaque AU, tandis que la séquence vidéo permet d'en capter la dynamique temporelle.

De même, les conditions d'illumination sont changeantes, et des visages partiellement masqués sont représentés également (lunettes, moustache, barbe, foulard,...).

Les prises d'images incluront au minimum les vues de face et de profil, et chaque Action Unit de la codification FACS doit être illustrée par l'ensemble de ses expressions.

Database	No. of Subjects	Elicitation	Imaging	Camera View	Labels
AR [85]	126	Posed	Static	Frontal Occlusion	Emotions
Belfast	125	Interviews and TV	Video	Frontal	Emotion and dimensions
Cohn-Kanade [58]	97	Posed	Video	Frontal	FACS AU
Cohn-Kanade+ [76]	123	Posed and Conversation	Video	Frontal and 15° to the side	FACS AU Emotion Landmarks
FABO	23	Posed	Video	Frontal	Emotion
GEMEP	10	Acted	Video	Frontal	Emotion
KDEF	70	Posed	Static	Five views	
JAFFE [81]	10	Posed	Static	Frontal	Emotion
MMI [82, 122]	101	Posed Spontaneous	Static Video (5 min)	Frontal 90° to the side	FACS AU
Face Database MPI [99]		Posed	Video	11 views at 18° intervals	FACS AU
Multi-PIE [48]	337	Posed	Static	15 views 19 illuminations	Emotion Landmarks
Prkachin-Solomon Pain [78]	129	Pain induction	Video	Frontal	AU Landmarks
Multi-PIE [64]	72	Posed	Static	Five views	Emotion
RU-FACS [7]	100	Interview	Video (2 min)	Mostly frontal	FACS AU
University of Texas Video Database [91]	284	Viewing videoclip	Video (10 minutes)	Frontal	Emotion
Bosphorous	105	Posed	Static	3D	FACS AU Emotion
BU-3DFE [130]	100	Posed	Static	3D	Emotion
BU-4DFE	101	Posed	Dynamic	3D	Emotion

FIGURE 3.1.1 – Comparaison des principales bases de données ouvertes.

Source : De la Torre, F. & Cohn, J. F. (2011). Facial Expression Analysis. In T. B. Moeslund, A. Hilton, V. Krüger, & L. Sigal (éd.), Visual Analysis of Humans : Looking at People (p. 377-409).

3.2.2 Sémantique de l'expression en termes d'affects

La sémantique n'est pas la même selon qu'une image est posée ou exprimée de manière spontanée. On va relever deux facteurs qui expliquent ceci. D'abord, les dynamiques temporelles de deux expressions simulées et spontanées, ne sont pas les mêmes.

Ensuite, il existe certaines AUs qui ne sont pas imitables aisément par une part importante de la population, mais qui par contre, sont fréquemment exprimées de manière spontanée [Pantic et al., 2005].

Également, les individus n'ont pas tous les même capacités à imiter certaines expressions.

L'idéal serait donc de pourvoir combiner pour chaque AU des images posées et spontanées.

C'est le but que se sont fixé les chercheurs avec la première base de données d'expression authentiques [Sebe et al., 2007], avec les contraintes suivantes destinées à éliminer tout biais :
a) le sujet ne peut pas être conscient qu'il participe à une expérimentation,
b) chaque sujet fait l'objet d'une interview après la prise d'image, pour rapporter son vécu émotionnel authentique,
c) rendre la présence de l'équipe de recherche invisible pour éviter tout biais.

Concrètement, cette expérimentation a été menée en amenant les sujets à interagir avec une borne interactive munie de caméras. Des extraits de film sont utilisés pour éliciter les émotions désirées.

3.2.3 Codage des méta-données

Ce sont les méta-données associées à chaque objet du dataset qui matérialisent cette vérité de terrain. Ces méta-données doivent au minimum inclure le tagging des AUs identifiées et leur dynamique temporelle.

Pour alimenter des algorithmes de reconnaissance de l'émotion efficaces, il est nécessaire de disposer de bases de données représentant les expressions et émotions correctement labellisées.
Le codage de ces méta-données est essentiel puisque sur lui repose la vérité de terrain sur laquelle seront entraînés et testées les modèles de classification.

Pour la base de données Cohn-Kanade+, la labellisation des émotions se fait en 3 passes :

1. Dans une première approche, les AUs détectés sur l'image sont comparés de manière stricte à la table de prédiction du dictionnaire FACS. Il en résulte qu'un label est donné si et seulement si tous les AUs qui le composent sont détectés dans l'image, à l'exclusion de tout autre.
2. Lors d'une deuxième passe, une comparaison plus lâche est faite : si dans une séquence apparaît un AU étranger à un label, on évalue si cet AU est cohérent ou non avec ce label. (Par exemple, l'AU 4, à composante négative, est cohérent avec le label de dégoût, mais pas avec celui de la surprise. Une extrait de la table de correspondance est illustré à la Fig.3.2.1.
3. La troisième étape est un jugement subjectif sur la ressemblance de l'expression affichée avec le l'élicitation demandée.

Le principal inconvénient est que le codage manuel est chronophage et coûteux. On estime [Pantic et al., 2005] le temps de codage par un codeur expert, à une heure pour 100 images. En même temps, il est recommandé que la même séquence d'images soit analysée par plusieurs codeurs en parallèle [Tian et al., 2005].

En ce qui concerne la *dynamique temporelle*, le déroulement d'une expression est caractérisée par 3 phases (voir fig.3.2.2), l'onset (qui suit l'expression neutre), l'apex et l'offset (qui précède le

Emotion	Criteria
Angry	AU23 and AU24 must be present in the AU combination
Disgust	Either AU9 or AU10 must be present
Fear	AU combination of AU1+2+4 must be present, unless AU5 is of intensity E then AU4 can be absent
Happy	AU12 must be present
Sadness	Either AU1+4+15 or 11 must be present. An exception is AU6+15
Surprise	Either AU1+2 or 5 must be present and the intensity of AU5 must not be stronger than B
Contempt	AU14 must be present (either unilateral or bilateral)

FIGURE 3.2.1 – Table de correspondance présence et absence d'AUs par label émotionnel.

Source : [Lucey et al., 2010]

retour à une expression neutre). C'est le marquage temporel entre les durées de ces 3 phases qui permet d'évaluer par exemple le caractère spontané ou non d'une expression, ou son intensité.

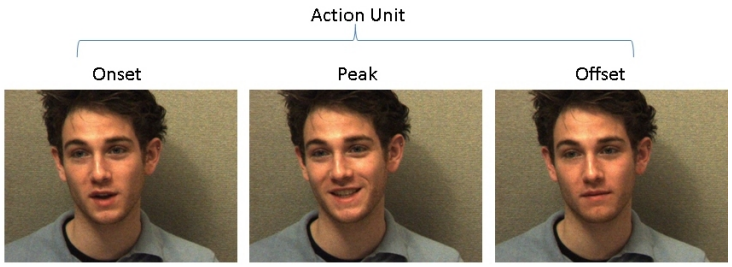


FIGURE 3.2.2 – Le phasage temporel de l'Action Unit 12.

Source : De la Torre, F., & Cohn, J. F. (2011). Facial Expression Analysis. In T. B. Moeslund, A. Hilton, V. Krüger, & L. Sigal (éd.), Visual Analysis of Humans : Looking at People (p. 377-409).

3.2.4 Diversité socio-démographique

Il est nécessaire de représenter un échantillon varié sur le plan du genre, de l'âge, de le diversité ethnique, des accessoires cosmétiques ou du style. Dans ce domaine également, un certain nombre de datasets ont été conçus de manière trop locale ou trop expérimentale.

Par exemple : Japanese Female Facial Expression (JAFFE)[Lyons et al., 1998] est basé sur des profils de femmes japonaises, MMI [Pantic et al., 2005] ne recense que des sujets de peau blanche, entre 20 et 33 ans, [Sebe et al., 2007] est centré sur un faible nombre de sujets issus de la population étudiante en majorité blanche.

3.2.5 Quantité de données

La taille de l'échantillon est enfin un paramètre limitatif. La précision de la prédiction de modèles augmente avec le nombre d'images disponibles pour les entraîner.

Pour obtenir un système de reconnaissance robuste sur les seules émotions prototypiques, on estime que 5.000 à 10.000 images sont nécessaires pour chaque AU [Lucey et al., 2010].
Or, les datasets actuels comptent plutôt de l'ordre de quelques centaines à quelque milliers d'images toutes expressions confondues.

Capturer plus de données en environnement naturel semble une option logique pour renforcer l'efficacité des algorithmes de classification. Pour illustrer la difficulté à collecter du matériel exploitable, l'expérience menée par Affectiva en 2013 (encore spin-off du MIT à cette époque) fournit des chiffres intéressants.

L'objectif était de constituer un dataset d'émotion spontanées, capturées auprès d'internautes via un plugin flash activant leur webcam.

— 6729 vidéos ont été capturées, dont 3268 exploitables,
— 489 internautes ont marqué leur accord pour que leurs images fassent l'objet de recherches,
— 242 vidéos ont été sélectionnées et codées manuellement, représentant 168.000 images,
— chaque image a été codée par minimum 3 codeurs différents issus d'un pool de 16 codeurs,
— le codage comprend 14 Action Units sur les 44 et 2 mouvements de tête.

Cet exemple est illustratif de la difficulté à collecter des images, tant sur plan de la capture technique, de l'obtention de l'accord du sujet, que des ressources à engager pour obtenir un codage partiel d'un échantillon de faible taille.

En conclusion de cette section, on note que les principaux datasets sont de taille assez réduite, qu'ils sont peu représentatifs et qu'ils recourent encore largement aux images posées.

Sur base de ces paramètres, la base de donnée CK+ [Lucey et al., 2010], qui est l'extension récente de la base de données Cohn-Kanade citée plus haut, reste le benchmark le plus souvent cité, avec la meilleure représentativité socio-démographique et la codification la plus complète (des AU et des labels d'émotions).

Il est important de préciser ici que la comparaison illustrée à la Fig.3.1.1 reprend uniquement les bases de données publiques, utilisées en recherche. Par comparaison, la startup Affectiva annonce disposer de 5,3 millions de vidéos,[1] de sujets provenant de 75 pays, tous usages confondus. Cette base de données étant privée, il n'est pas précisé combien servent à la reconnaissance de l'émotion, ni selon quels protocoles elles ont été collectées (par leurs partenaires commerciaux).

1. Gabi Zijderveld, Affectiva, http ://blog.affectiva.com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting

Dans le cadre de la recherche scientifique, on peut néanmoins considérer que la taille des échantillons d'images pose la question de la fiabilité de la reconnaissance de l'émotion en temps réel.

Nous y reviendrons dans les conclusions de ce travail.

3.3 Fiabilité de la base expérimentale

En dépit de la popularité des FACS, il existe peu de littérature ayant trait à la fiabilité en général pour les expressions posées, et à fortiori pour les expressions spontanées, considérées comme moins fiables encore. La plupart des études sur le sujet fournissent des chiffres prenant en compte la moyenne de l'ensemble des AUs mesurées [Cohn et al., 2007].

Nous interrogeons ci-après la fiabilité des FACS à deux niveaux : a) quelle serait la fiabilité de la codification FACS en tant que telle, c'est-à-dire le taux d'accord inter-juges entre codeurs FACS en laboratoire et b) quelle serait le taux d'accord entre cette codification manuelle et celle que peut atteindre un système automatique de détection ?

Concernant la fiabilité du codage humain, la source la plus exhaustive sur le sujet est constituée par l'évaluation de Sayette *et.al* [Sayette et al., 2001], qui propose comme unité de mesure le taux d'accord exprimé par son coefficient kappa.

La prise en compte du taux d'accord inter-juges est proposé sur les critères de validité suivants et mesurée au cours d'une triple expérimentation en laboratoire :

- La validité pour chaque AU individuellement,
- Le taux de précision temporelle : il s'agit de la fenêtre minimum de temps dans laquelle l'occurrence d'une action faciale est correctement codée. Partant d'une fenêtre d'1/30 seconde, jusqu'à 1/6 et 1/2 seconde, on évalue l'augmentation du taux d'accord. C'est une mesure de la capacité à localiser correctement le commencement d'une AU dans le temps.
- L'intensité : 5 AUs seulement sont prises en compte pour la mesure de l'intensité. Il n'existe pas de base expérimentale pour le coefficient kappa de mesure de l'intensité entre juges. Le processus de certification pour accéder au titre de codeur FACS ne le mentionne pas.
- La corrélation entre AUs ou combinaisons d'AUs et les 6 émotions discrètes (joie, surprise, tristesse, colère, peur, dégoût), plus une valeur plus abstraite d'émotion positive/négative.

On peut résumer ainsi ce qu'il ressort de l'expérimentation :

1. Les taux d'accord pour la précision de la détection et l'intensité sont bons à excellents.
2. La précision temporelle augmente le plus fortement entre 1/30 et 1/6 de seconde, indiquant plutôt la difficulté du codeur à identifier avec précision le moment où localiser le début de l'action. Ce paramètre est typiquement lié à l'expérience du codeur.
3. La corrélation émotion / Action Units est bonne à excellente.

Cette étude, bien qu'une des plus complètes publiées sur le sujet, a ses limites :

- L'étude a été menée dans un seul laboratoire et ne préjuge pas de possibles différences de traitement entre équipes différentes,
- la validité des FACS comme mesure de l'émotion, c'est-à-dire la validité du lien universel entre AUs et émotion, ne fait, au jour de la publication, pas l'objet d'évaluation (nous n'avons pas de source ultérieure à faire valoir),
- 30 AU's ont été prises en compte sur les 44, les autres présentant un nombre d'occurrences trop faible.
- La corrélation aux émotions est mesurée pour les seuls labels dégoût et tristesse, ainsi que positive/négative.

En conclusion, la valeur probante intrinsèque de cette étude est limitée. En même temps, la succès des théories d'Ekman vient précisément du fait qu'il a su, en collaboration avec le domaine informatique, imposer les FACS comme standard unique, dont la fiabilité est rarement questionnée.

3.4 Codage automatique des actions faciales

Le second niveau auquel on s'intéresse est la fiabilité avec laquelle une système de détection automatique peut réaliser la même tâche que l'humain, à savoir réaliser la détection des Action Units individuelles ou en combinaison, et leur dynamique temporelle.

Ce champ d'investigation est très transversal, car capter des informations de bas niveau est nécessaire dans un grand nombre de domaines, comme l'étude des maladies mentales, l'étude de la concentration, des techniques d'apprentissage, etc.

Cela questionne également s'il est possible par les techniques d'analyse d'image, de corriger les biais liés à l'analyse humaine : le jugement humain est en partie subjectif, non standardisé d'un laboratoire à l'autre, et pas toujours efficace pour localiser les bornes temporelles d'une action.

Valstar et al. [Valstar and Pantic, 2012] (Fig.3.4.1)proposent une implémentation d'un système permettant de capturer 22 Action Units ainsi que leur développement temporel.
Les tests démontrent que le système est capable de détecter les AUs avec un haut degré de précision, et qu'il est de plus bien généralisable, c'est à dire qu'entraîné sur un dataset et testé sur un autre, il se comporte toujours bien.
En revanche, le système ne fonctionne pas sur des données totalement nouvelles (unseen data), ni réellement en environnement réel, puisqu'il échoue à partir de 20 degrés d'angle de rotation de la tête.

Dans une autre contribution, la fiabilité d'un système automatique de classification des AUs par rapport à la détection manuelle a été évaluée [Cohn et al., 1999]. A l'aide d'un classifieur utilisant l'analyse discriminante, les images d'un échantillon de 10 étudiants de 18 à 30 ans sont analysées et les résultats des 15 AUs les plus fréquentes sont comparées au travail de deux codeurs certifiés FACS.

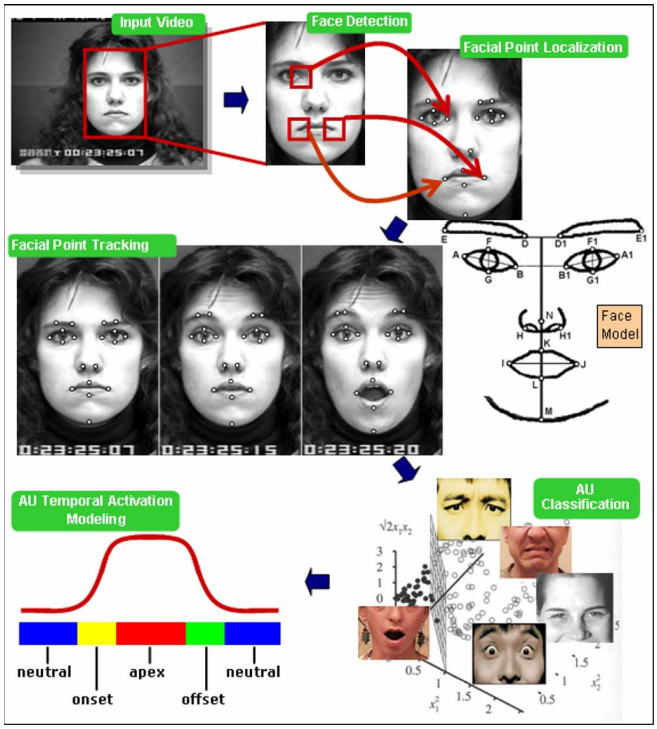


FIGURE 3.4.1 – La détection automatique de 22 AUs et de leur dynamique temporelle : vue générale.

Source : [Valstar and Pantic, 2012]

Les taux d'accord se montent à 81%, 88% et 91% respectivement pour les AUs liées aux yeux, à la bouche et aux sourcils, et 82% pour l'identification du sourire de Duchenne par rapport à un sourire feint. Le niveau d'accord inter-observateur est jugé en ligne avec celui atteint par codage manuel selon le standard FACS.

Il faut noter que le système requiert encore la calibration manuelle de l'image de début de chaque séquence, ce qui a représenté 4 heures de travail pour un codeur certifié, pour marquer 504 images. En comparaison, 10 heures de travail par minute de séquence sont nécessaires pour le codage manuel [Ekman, 1982].

3.5 Conclusion

La constitution d'une base de données d'images est nécessaire pour l'entraînement, le test, la validation de tout système de reconnaissance automatique, ainsi que comme benchmark pour la comparaison inter-systèmes.

Les premiers datasets ont été constitués sur base du corpus théorique de la théorie des émotions universellement interprétables de Paul Ekman, et sa codification sous forme de FACS.

Constitués à l'origine d'images statiques illustrant les 6 émotions de base sur un mode posé, les plus récentes intègrent maintenant, pour tout ou partie, des images spontanées, ce qui permet d'améliorer la généralisabilité d'un système en le rendant capable de reconnaître un éventail d'expressions plus variées.

La vérité de terrain sur laquelle repose l'évaluation d'un système de reconnaissance automatique de l'expression est constituée par les méta-données adjointes aux images :

— Action Units marquées par observation manuelle, image par image de la séquence, par le protocole de codification FACS pour l'immense majorité.

— Labels émotionnels pour chaque séquence, par jugement perceptuel plus subjectif,

— Mesure de la temporalité précise entre l'onset, l'apex et l'offset,

— Mesure de l'intensité (sur une échelle de 5 niveaux pour le FACS).

A ce jour, la constitution d'un dataset reconnu comme benchmark fait défaut, tant pour la calibration des modèles de classification, que pour la mise en place d'un standard de mesure de performance entre différents systèmes.

On peut encore citer les limitations suivantes dans le domaine de la construction de bases de données :

- La taille des échantillons reste trop faible pour obtenir des modèles de prédictions robuste sur toutes les expressions faciales que peut afficher un isage de manière spontanée,
- La grande majorité des bases de données sont structurées autour de la théorie des émotions universelles, limitant le nombre d'émotions aux 6 émotions de base,
- Le codage manuel, incluant la détection des AUs, le jugement perceptuel d'un label émotionnel, l'intensité et la temporalité est toujours lent et coûteux,
- S'il existe un protocole de codage et de mesure de l'accord inter-juges bien spécifié dans le FACS, il n'existe pas de spécification quant à la représentativité de ce que devrait être un échantillon optimal, en termes d'ethnicité, de genre, d'âge ou de contexte culturel.

Chapitre 4

Une alternative à la théorie des émotions discrètes : l'Appraisal Framework

4.1 Introduction

Nous avons vu que l'approche théorique la plus largement appliquée en informatique pour la reconnaissance de l'émotion découle de la théorie des émotions discrètes issue des travaux de Paul Ekman.

On peut considérer qu'il s'agit d'une approche pragmatique, puisqu'elle évite les controverses techniques sur la nature même des émotions, telles que nous les avons évoquées dans le premier chapitre.

6 émotions universelles reposant sur des configurations faciales prototypiques combinées à la nomenclature FACS font de la théorie des émotions de base le corpus conceptuel le plus à-même d'être traduit en langage informatique.

Pour autant, l'état de l'art ne propose encore qu'une réponse parcellaire au problème de la reconnaissance faciale de l'émotion en conditions réelles. On peut résumer les limitation actuelles en 5 points [Calvo and D'Mello, 2010], sur base de l'état de l'art produit en [Zeng et al., 2009] :

1. Le set d'émotions détectées se limite pour la plupart à 6 émotions universelles,
2. L'entraînement des modèles est encore souvent basé sur des images posées, en environnement contrôlé,
3. Un cinquième des solutions n'opèrent pas en real-time,
4. La majorité requièrent une pré-segmentation manuelle des images (tagging des frames d'apex, et état neutre),
5. Enfin presque tous sont insensibles au contexte, là où l'expérience émotionnelle est préci-

sément « un événement riche de contexte »[1].

Un autre point de vue est qu'une approche pragmatique ne peut pas faire l'impasse sur un modèle théorique. Les recherches en sciences informatiques ne peuvent pas être découplées des sciences affectives et de l'héritage des travaux en psychologie sur la compréhension de l'émotion.

Dans ce chapitre, nous présentons un modèle théorique alternatif au modèle des émotions universelles [Mortillaro et al., 2012], basé sur la théorie de l'évaluation cognitive. Il ne s'agit d'une perspective théorique issue du champ des sciences affectives, mais il nous paraît pertinent de la mentionner dans ce travail, dans la mesure où elle apporte à la fois une critique pertinente du modèle discret de l'émotion et une base conceptuelles pour élargir le champ d'un système de reconnaissance de l'émotion à une plus large palette de nuances d'affects.

4.2 Comparaison des différents modèles de l'émotion

Dans ce chapitre, nous présentons le modèle de l'appraisal (évaluation cognitive), qui se pose en synthèse entre les deux modèles discrets et continus. Nous rappelons ci-dessous brièvement les 3 modèles, et quelles forces et faiblesses il présentent pour être traduits en systèmes informatiques.

4.2.1 Modèles discrets

Les modèles de reconnaissance de de l'émotion basés sur l'approche discrète, sont extrêmement efficaces sur la détection des émotions de base, à partir d'expressions posées et non-spontanées. (Typiquement celles que l'on retrouve dans les bases de données servant à l'apprentissage des systèmes, comme Cohn-Kanade [Kanade et al., 2000] ou JACFEE [Biehl et al., 1997]).

La question qui est posée est celle de savoir quelle est la validité de ce que l'on infère à partir d'une expression : est-ce une émotion ou de manière plus rudimentaire, une expression du visage, n'ayant pas la même signification affective.

Trois critiques sont formulées :

1. la conclusion à un label émotionnel à partir d'une configuration dAUs repose sur la seule théorie des émotions discrètes,
2. les émotions spontanées sont largement moins bien interprétées,
3. la détection est limitée aux 6 émotions de base.

4.2.2 Modèles dimensionnels

En décomposant l'espace émotionnel en un champ à 2 (ou 3) dimensions, ce modèle a pour avantage de simplifier le processus de reconnaissance en le limitant à un couple valence/activation pour l'attribution d'un label.

1. Barrett, Lisa Feldman, et al. "The experience of emotion." Annu. Rev. Psychol. 58 (2007) : 373-403

Mais de la même manière que dans le modèle discret, ne disposer que de 2 ou 3 variables paraît limitatif pour saisir la complexité d'une émotion. Ainsi par exemple, la peur et la colère, toutes deux qualifiées par un faible niveau de valence et un haut niveau d'activation, ne sont pas aisément distinguables.

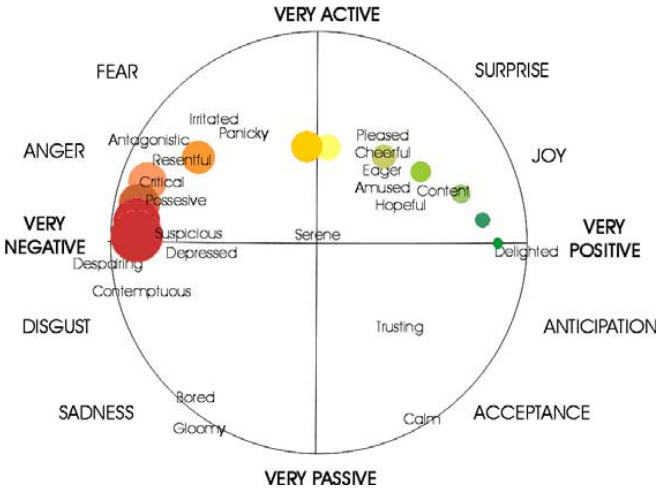


FIGURE 4.2.1 – Positionnement des émotions sur le modèle dimensionnel.

Source : [Lucey et al., 2010]

De manière générale, l'utilisation de dimensions continues ne permet pas de spécifier clairement les frontières entre les différentes émotions.

Une solution est de réduire le raisonnement à deux décisions binaires de type valence basse/élevée et activation faible/forte, mais avec comme conséquence de perdre l'avantage de la représentation plus subtile qu'a le modèle dimensionnel sur le modèle des émotions de base.

4.2.3 Modèles de l'évaluation cognitive

Le modèle de l'appraisal emprunte aux deux courants : de la théorie des émotions de base que chaque émotion est porteuse d'états subjectifs distincts, et du modèle dimensionnel qu'elle résulte de plusieurs dimensions sous-jacentes.

Des 3 modèles, le modèle de l'appraisal est le plus complexe, car il fait entrer en ligne de compte les différences propres à l'individu et au contexte. Un même stimuli pourra déclencher l'apparition d'émotions différentes chez une même personnes, à des moments différents. Le modèle ne postule pas de relation 1-1 entre une situation et une réponse émotionnelle, ni entre un appraisal particulier et un label émotionnel.

Comme illustré dans la figure 4.2.2, tout stimulus est évalué par rapport à sa signification et ses conséquences pour le sujet. C'est ce processus d'évaluation qui consiste à attribuer des

valeurs distinctes à ces variables d'évaluation, la résultante étant un vecteur de variables pour lequel une correspondance avec une émotion est possible.

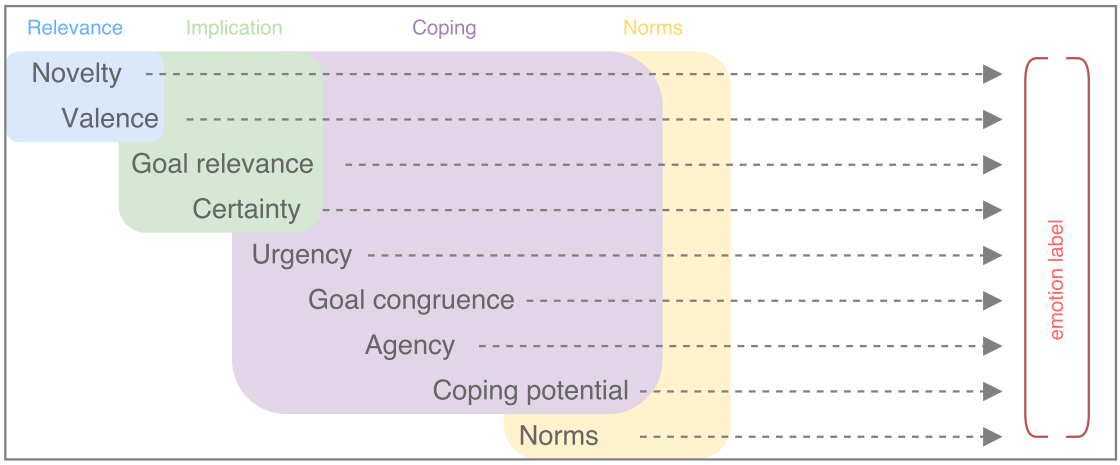


FIGURE 4.2.2 – Les variables d'appraisal (représentation simplifiée)

Source : adapté de Hudlicka, E. Computational Analytical Framework for Affective Modeling : Towards Guidelines for Designing Computational Models of Emotions. IGI Global.

Le processus d'identification d'une émotion correspond dès lors à comparer ce premier vecteur aux valeurs décrites dans un tableau de correspondance illustré à la Fig.4.2.3.

Ce tableau fait le lien entre les variables contenues dans le vecteur de la Fig.4.2.2 et les labels émotionnels.

4.3 Application à la reconnaissance des émotions : le Component Process Model

Le modèle des émotions discrètes propose comme représentation 6 labels, le modèle continu propose le circumplex (Fig.4.2.1, le modèle de l'appraisal a défini le CPM comme modèle de représentation.

Le modèle de l'appraisal spécifie une séquence d'évaluation subjectives, débouchant sur une séquence de variables représentant la signification personnelle et subjective de l'événement pour le sujet. C'est ce processus subjectif, contextuel et individuel qui est à la source des réactions physiologiques et expressives. Des patterns d'expression sont spécifiés, qui permettent dès lors d'associer un vecteur de variables d'appraisal à une expression, spécifiée sous forme d'AUs. (voir Fig.4.3.1).

la spécificité des théories de l'appraisal est qu'elles font des hypothèses sur le contexte indivi-

Appraisal Variable	Fear	Anger	Joy	Sadness	Shame	Guilt	Pride
Relevance							
Novelty							
Suddenness	HIGH	HIGH	HIGH/ MED	LOW	LOW	open	open
Familiarity	LOW	LOW	open	LOW	open	open	open
Predictability	LOW	LOW	LOW	open	open	open	open
Valence	LOW	open	open	open	open	open	open
Goal Relevance	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
Implications							
Cause: Agent	OTHER/NAT	OTHER	open	open	self	self	self
Cause: Motive	open	INT	INT/ CHAN	INT/CHAN	INT/ NEGLIG.	INT	INT
Outcome Probability	HIGH	V. HIGH	V. HIGH	V. HIGH	V. HIGH	V. HIGH	V. HIGH
Discrepancy from Expectation	DISS	DISS	open	open	open	open	open
Conduciveness to Goal	OBSTR	OBSTR	V. HIGH	OBSTR	open	HIGH	HIGH
Urgency	V. HIGH	HIGH	LOW	LOW	HIGH	MED	LOW
Coping Potential							
Control	OPEN	HIGH	open	V. LOW	open	open	open
Power	V. LOW	HIGH	open	V. LOW	open	open	open
Adjustment	LOW	HIGH	MED	MED	MED	MED	HIGH
Normative Significance							
Internal Standards	open	open	open	open	V,LOW	V.LOW	V.HIGH
External Standards	open	LOW	open	open	open	V.LOW	HIGH

FIGURE 4.2.3 – Tableau de correspondance entre les variables d'appraisal les émotions (extrait).

Source : tiré de Hudlicka, E. Computational Analytical Framework for Affective Modeling : Towards Guidelines for Designing Computational Models of Emotions. IGI Global.
Le terme "open" signifie qu'il peut exister plusieurs résultats d'évaluation correspondant au label émotionnel donné, ou que le valeur n'est pas significative.

duel et culturel du sujet, tenant compte des différences entre les réponses émotionnelles de deux personnes à un même événement. Le contexte culturel est pris en compte également : sur base d'expérimentations, il est possible d'établir des différences dans les émotions ressenties entre les cultures occidentale individualiste et japonaise plus collectiviste.

Les américains par exemple attribuent plus facilement la causalité à eux-mêmes en cas de succès, (ressentant la fierté) ou aux autres en cas d'échec (ressentant la colère), que les japonais, qui les évaluent plutôt comme la chance ou la culpabilité [Moors et al., 2013].

4.3.1 Le Component Process Model

Le CPM postule que l'émotion est majoritairement définie par la composante cognitive, qui définit une série d'évaluations, les SEC, résultant en un espace multi-dimensionnel de variables, qui permettent par la suite d'inférer des prédictions.

Le CPM produit des prédictions sur l'expression faciale, les mouvements corporels et la voix. La figure 4.3.1 illustre les prédictions sur l'expression faciale, la voix et le corps. Ces prédictions sont basées sur un corpus de résultats empiriques issus de l'observation des sujets et des considérations physiologiques.

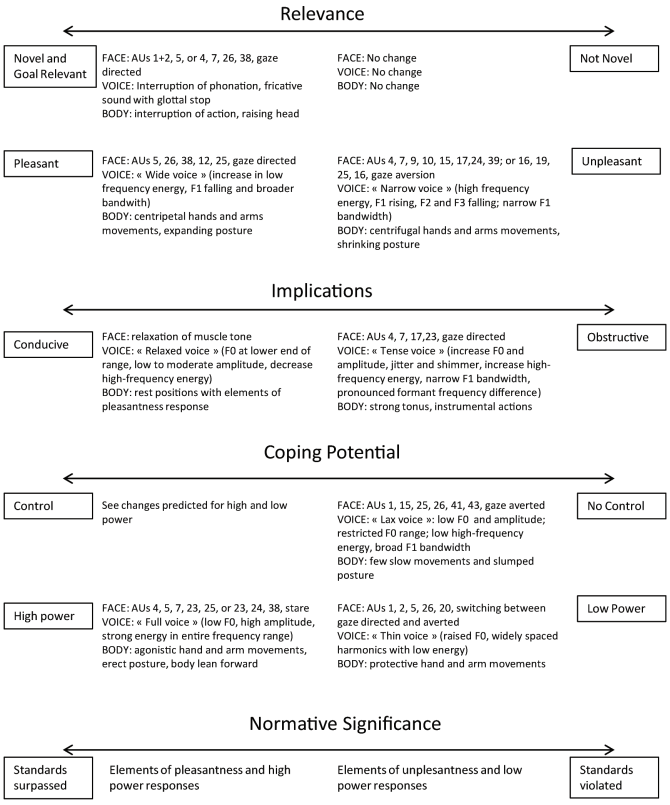


FIGURE 4.3.1 – prédictions du Component Process Model. (AU : "Action Unit", from Ekman &Friesen, [Ekman and Friesen, 1978])

Source : Mortillaro, M. et.al (2012). Advocating a Componential Appraisal Model to Guide Emotion Recognition. International Journal of Synthetic Emotions (IJSE), 3(1), 18-32.

Ici, un lien est fait avec le domaine de la modélisation des émotions, qui est très utilisée dans le champ de l'*Affective Computing*. On peut définir l'*Affective Computing* comme « *l'étude et le développement de systèmes et équipements capables de reconnaître, interpréter, traiter et simuler les affects humains* »[2].

L'adaptation du modèle à la reconnaissance de l'émotion est proposée de la manière suivante.

2. Picard, Rosalind Wright, et al. Affective computing. 1995.

En inversant le processus, il est posé que les variables d'appraisal peuvent être inférées à partir de l'expression faciale.

L'idée repose sur l'hypothèse, testée en [Scherer and Grandjean, 2008], qu'à partir d'images représentant des expressions émotionnelles, lorsqu'il est demandé à des observateurs d'en faire la description par

a) des labels d'émotions,

b) des appraisals,

c) des interactions sociales,

d) des tendances à l'action, c'est par la description des appraisals que le taux d'accord inter-observateur est le meilleur.

L'utilisation du modèle de l'appraisal répond à 3 besoins :

Premièrement, établir un lien entre reconnaissance et synthèse d'émotions. « *La capacité d'un système à détecter les états affectifs dépend de sa capacité à en identifier les patterns d'expression* »[Picard, 1995].

L'utilisation des variables d'appraisal constitue ici une interface qui permet d'assurer la compatibilité entre les 2 modèles de reconnaissance et de production de l'émotion.

Dans le modèle discret (voir fig. 4.3.2, cadrans du haut), la détection est basée sur les expressions faciales exprimées sous forme de FAU et un label émotionnel est généré, sans qu'il y ait modélisation des facteurs sous-jacents qui produisent cette émotion.

En sens inverse, il est posé que l'évaluation d'un événement, selon un « programme affectif »produit telle ou telle émotion, mais sans en appréhender les différentes composantes individuellement.

Avec le modèle du CPM, il s'agit de rendre ce lien explicite, ce qui n'est pas fait dans la théorie des émotions discrètes. Si dans la synthèse d'émotions, les individus procèdent à des évaluations cognitives et il en résulte des ressentis émotionnels, alors il apparaît plausible que des observateurs, puissent, à partir des résultantes de ces appraisals (des AUs par exemple), inférer une émotion à partir de ces variables d'appraisal.

Ce lien entre synthèse et émotion est illustré ci-dessous (Fig. 4.3.2, cadran du bas).

Pour les auteurs, l'avantage d'un système de détection qui génère non plus des labels d'émotion de base, mais des vecteurs de variables d'appraisal est à même d'apporter une détection plus fine dans des cas où les émotions sont subtiles, diverses ou bien lorsqu'aucun label ne peut être directement associé à une configuration d'AU.

Ensuite, la prise en compte du contexte du sujet dans l'expérience est également un aspect important, bien qu'encore peu expérimenté [Castellano et al., 2010].

On peut citer par exemple [Pantic and Rothkrantz, 2004], qui implémente un système de classification des expressions qui fait son apprentissage sur base d'une mémoire dynamique enrichie par le jugement au cas par cas d'un utilisateur.

Ce système permet en l'occurrence d'ajouter aux 6 émotions de base, des états comme « ennuyé »ou « enchanté ».

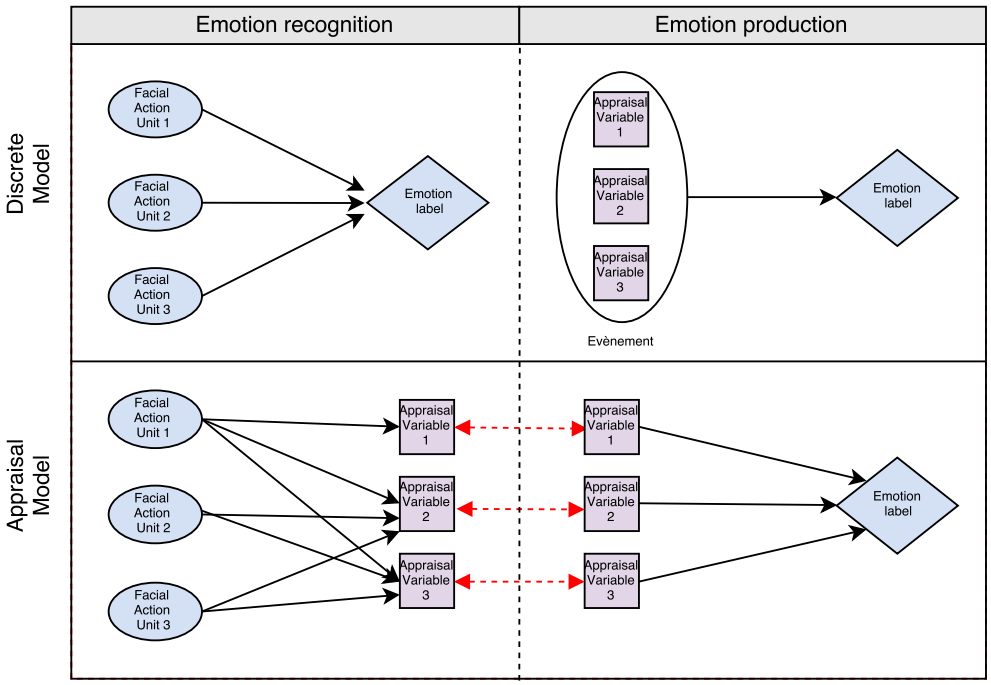


FIGURE 4.3.2 – Comparaison du lien entre détection et génération d'un label d'émotion.

Source : adapté de [Mortillaro et al., 2012].

Les évaluations cognitives représentent en réalité une abstraction du contexte en tant que tel. Il devrait donc être possible, si l'on peut inférer cette évaluation cognitive des expressions, d'en déduire non seulement les labels émotionnels, mais également les causes qui les sous-tendent.

Le modèle de l'appraisal est largement majoritaire dans les modèles de synthèse de l'émotion, car il est le seul qui spécifie totalement le lien entre cognition et production de l'émotion [Marsella et al., 2010], d'où sa pertinence pour la conception d'agents intelligents dans la robotique par exemple, ou d'Human-Computer Interaction (HCI) capables d'interpréter l'affect de l'utilisateur et d'en produire en retour de

Enfin, un système de reconnaissance d'émotion utilisable dans le monde réel doit être capable d'aller plus loin que les émotions prototypiques (émotions de base).

Dans la vie réelle, ces émotions de base constituent plus souvent l'exception que la règle. Par exemple, un léger haussement de sourcil seul ne peut être assimilé à une émotion pleine, mais plutôt au caractère inattendu d'une situation.

L'adaptation des méthodes de reconnaissance aux comportements d'utilisateurs réels nécessite la prise en compte de ces états émotionnels « subtils » [Castellano et al., 2010] [Zeng et al., 2009].

Le CPM produit des prédictions pour sur les émotions de honte/fierté, culpabilité, ennui, désespoir, anxiété, colère intense, qui n'existent pas dans le modèle des émotions discrètes.

4.4 Conclusion

Le panorama actuel de la recherche sur la détection de l'émotion est totalement dominé par la théorie des émotions de base, ou émotions discrètes.

Si cette dernière adopte un point de vue pragmatique, faisant l'impasse sur les questionnements théoriques, elle porte plusieurs limitations : validité émotionnelle reposant sur la seule théorie des émotions universelles, nombre réduit d'émotions prises en compte et difficulté à appréhender les expressions mixtes ou subtiles.

Dans ce chapitre, nous avons trouvé utile de présenter le CPM comme modèle alternatif de représentation de l'émotion à haut niveau, en réponse aux li

Le CPM repose sur la théorie de l'évaluation cognitive, qui est la plus utilisée dans le domaine de la recherche sur la synthèse d'émotion. Il apparaît donc possible d'inverser le processus pour le rendre pertinent dans la détection des émotions.

Dans la phase de détection, des inputs (qui peuvent être sous forme d'Action Units dans le cas de la reconnaissance faciale) peuvent être traduits selon la table de prédiction du CPM (Fig. 4.3.1), en une série d'appraisals qui sont propres à l'individu, compte tenu de son vécu, sa culture et dans son contexte.
Ces appraisals servent d'interface à partir desquelles il est ensuite possible d'inférer un label émotionnel, sur base d'une spécification basée sur l'étude empirique (Fig.4.2.2).

Par rapport à la théorie des émotions universelles, les avantages sont les suivants ; premièrement le modèle de l'appraisal postule un plus grand nombre d'émotions, et une granularité plus fine des émotions mixtes. Ensuite il rend possible la prise en compte du contexte en établissant un lien explicite, par la nature même des appraisals, entre le sujet et son environnement.

L'universalité de l'émotion est traitée différemment dans les théories de l'appraisal. Le contexte culturel est pris en compte dans le processus d'évaluation. Face à une même situation, des sujets japonais et américains ressentent des émotions différentes en situation de succès ou d'échec.

Enfin, un lien formel est spécifié entre synthèse et détection de l'émotion. Dans un premier temps, le sujet réalise une série d'évaluations à une situation dans un contexte, qui déclenchent une émotion.

Dans un second temps, un observateur va percevoir le résultat de ces évaluations (à travers des Action Units sur le visage par exemple) et en inférer un label émotionnel, couvrant un éventail d'émotions non identifiées par le modèle discret (honte/fierté, culpabilité, ennui, désespoir, anxiété, colère intense).

Chapitre 5

Conclusion

La technologie de reconnaissance faciale de l'émotion n'est pas qu'un artefact technique. Par son adaptabilité aux objets connectés largement répandus dans notre environnement (caméras de surveillance, smartphones, webcams, etc.), la reconnaissance de l'émotion a un potentiel à la fois inéluctable et invasif dans l'analyse et la prédiction de nos comportements.

Comme nous l'avons vu au chapitre II, 1a technologie hérite de fondements théoriques et force est de constater que les implémentations des systèmes reposent exclusivement sur la conception des émotions de base formulée par Silvan Tomkins et Paul Ekman.

Il existe une double dépendance à cette théorie : d'une part sur la représentation des émotions (les 6 émotions de base), d'autre part sur la méthodologie de travail de codification des expressions faciales (le FACS). Il est donc pertinent de s'interroger sur la solidité de ce corpus dominant.

Deuxièmement, un des arguments souvent mis en avant pour justifier le recours à la biométrie est la neutralité de la technologie. Les algorithmes n'ont pas de conscience et ne sont, en eux-mêmes porteurs d'aucun biais, ils sont donc amenés à jouer un rôle positif dans la société.

Un classifieur n'opère qu'en cherchant des relations entre des données qu'on lui fournit comme étant la réalité de terrain sur base de laquelle analyser un problème. Cette réalité de terrain contenue dans les bases de données, forgée sur base de choix subjectifs, peut, en revanche porter en elle de potentiels biais.

Cette dernière section est articulée comme suit. Nous développons une critique en 3 points : premièrement, quelle est la validité du corpus théorique sur lequels sont construits les systèmes de reconnaissance de l'émotion, ensuite, sont-ils neutres, et quel est le niveau de subjectivité que l'on peut accorder à leurs résultats ?

Enfin, nous nous pencherons sur les biais potentiellement induits par la construction des bases de données.

5.1 Régime de validité

Quelle est la fiabilité que l'on peut accorder au concept de reconnaissance faciale de l'émotion, et comment la mesurer ?

La fiabilité technique est mesurée à travers des indicateurs chiffrés objectifs, tels que le taux d'accord inter-juges pour le traitement par les humains, le taux de reconnaissance d'un algorithme mesuré en faisant tourner un classifieur sur un échantillon de test préalablement labellisé.

La mesure de la performance d'une technique se doit d'être objectivée, des initiatives ont été mises sur pied à ce titre par la communauté scientifique.

Le Facial Expression Recognition Challenge (FERA) [Valstar et al., 2011] s'est tenu à deux reprises déjà, fixant un cadre méthodologique pour confronter différents algorithmes sur un *training set* unique et un *test set* de données non vues, mis à disposition en dernière minute.

Dans ce travail, nous cherchons plutôt à évaluer la fiabilité du concept de reconnaissance de l'émotion à travers les indices de mouvements faciaux.

L'inférence d'une émotion à partir du visage repose-t-elle sur des postulats valides ? Quels sont les biais embarqués dans les choix posés pour les construire ?

5.1.1 Critique de la validité de la théorie des émotions de base

Ce qui donne un aspect « objectif »au modèle des émotions proposé par Ekman et en fait un concept attirant pour la modélisation informatique, c'est son lien à la théorie de l'évolution : les émotions sont déclarées universelles, et surtout génétiquement programmées en nous, de sorte qu'à tout moment il existe un lien codé « en dur »entre notre état affectif et les mouvements qu'affiche notre visage, même involontairement.

Une fois établie la représentation de ce modèle sous forme d'Action Units, le reste le reste est abordé comme un défi technique d'apprentissage par la machine de ces patterns physionomiques.

Or cette conception n'est pas unanimement partagée dans toutes les disciplines. La base théorique sur laquelle se fonde l'analyse automatique des émotions est « entâchée d'incertitudes »[Bjørnsten and Zacher Sørensen, 2017].

La culture occidentale considère que le visage est porteur 1) de notre identité et 2) de notre ressenti émotionnel.

Trois critiques peuvent être faites à cette encontre. Premièrement, la théorie des émotions de base repose sur des hypothèses scientifiquement fragiles.

Deuxièmement, l'analyse en temps réel de l'image, son découpage en séquences d'images statiques et la classification en catégories discrètes ne peuvent rendre compte fidèlement du caractère évolutif, « polyphonique »du développement d'une expression faciale.

Et troisièmement, l'analyse de l'expression faciale étant encore non contextuelle, son utilisation ne se prête pas encore aux fins de prévision et d'anticipation du comportement.

Les fondements scientifiques de la théorie des émotions universelles sont largement contes-

tés dans les travaux de Lisa Fieldman Barrett qui dirige l' *Interdisciplinary Affective Science Laboratory* de la Northeastern University.

Barrett conteste [Gendron et al., 2013] l'existence d'émotions universellement reconnues. Il existe selon elle un biais, induit par les expériences originales d'Ekman et Tomkins, et répété depuis dans les suivantes : la liste des 6 émotions de base a toujours été suggérée en amont. Le postulat de leur existence est posé avant l'expérimentation. Et toutes les expérimentations ont amené les sujets à se prononcer en fonction de cette liste finie.

En modifiant le protocole de sorte à tester différentes émotions à travers les cultures sans en suggérer la liste au départ invalide les conclusions d'Ekman.

Ensuite, Barrett conteste l'idée d'un bagage génétique « codé en dur »dans le cerveau. La signification à donner à une émotion tient autant du contexte, des signaux du corps, de la voix et de l'expérience, que du visage. Le visage ne peut, à lui seul porter le sens du ressenti réel de l'individu.

Sa critique vise tant Apple que les services de sécurité du contrôle aérien, qui « devraient savoir qu'un visage ne raconte pas toute l'histoire »[1].

5.2 Régime de justice

Un des arguments en faveur de la reconnaissance faciale, de manière générale, est son caractère neutre, objectif, insensible au contexte social, politique, culturel.

Pris de manière isolée, l'artefact lui-même peut être neutre, mais ses concepteurs ne peuvent l'être totalement. Des biais peuvent être embarqués involontairement de cette manière.

Ainsi par exemple, le National Institute of Standards and Technologies (NIST) opère depuis 2011 des benchmarks, sur base volontaire, des systèmes de reconnaissance faciale déployés dans les lieux publics aux états-unis. Du rapport de 2012 [Klare et al., 2012] ressort qu'ils « performent »nettement moins bien auprès des individus jeunes, des noirs et des femmes.

Un premier enseignement est qu'une meilleure représentativité socio-démographique des datasets influence de manière positive le taux de reconnaissance.

Ensuite, pour un individu d'une cohorte donnée, l'algorithme se comporte mieux s'il est entraîné sur un dataset d'images d'âge/ethnie de la même cohorte, et enfin, en conclusion, il est suggéré le recours à la *sélection dynamique d'identificateur de visages* : les opérateurs disposeraient de suites d'algorithmes multiples entraînés spécifiquement sur des classes de population distinctes.

On peut en déduire qu'à la construction, les programmeurs de tels systèmes y intègrent des éléments propres à leur culture : il se peut qu'ils mettent l'accent plus particulièrement sur

1. L. Fieldman-Barrett, What Faces Can't Tell Us, The New-York Times, 28 février 2014. https ://nyti.ms/2iy44tW.

certains traits, ou bien qu'ils y intègrent leur propre expérience de la lecture physionomiste du visage.

Une telle analyse n'existe pas au niveau de la reconnaissance de l'émotion. Il n'est pas prouvé que ces résultats y soient exportables, mais le problème peut se poser néanmoins d'une manière différente. Pour un algorithme d'identification, la tâche consiste à trouver des relations entre l'image *codée* d'un visage et une identité, ce qui, comme nous venons de le voir, comporte une part d'incertitude. En reconnaissance de l'émotion, il s'agit d'établir des relations entre une image *sémantique* et un label émotionnel, l'incertitude est de nature différente, car c'est cette image sémantique qui est incertaine, puisque reposant sur le jugement subjectif embarqué dans la labellisation des images d'un dataset.

5.3 Validité et biais induits par les bases de données

Un des principaux défis de long terme pour la reconnaissance faciale de l'émotion est d'atteindre la détection en temps réel et en conditions réelles.

Or la quantité de données disponibles dans le training set représente un facteur plus limitatif que la méthode de classification [Valstar et al., 2011].

Si l'on songe aux quelques 7000 configurations musculaires différentes affichables par un visage, plus les autres facteurs de variation (illumination, mouvements de la tête, occlusions,...), cela représente des quantités énormes d'images à codifier dans un training set idéal.

Nous identifions 3 facteurs pouvant influencer la fiabilité par les biais induits à l'origine par la constitution des datasets : la quantité de données à collecter, la part subjective du codage, et les biais inhérents au dataset.

La quantité de données à collecter pour le passage à l'échelle d'un set d'entraînement conforme à le vie réelle est trop importante pour que le codage puisse être fait manuellement. La différence avec un prototype de laboratoire entraîné sur quelques centaines de clichés par 1 ou 2 codeurs est dans le nombre d'images et le nombre de codeurs nécessaires.

Une solution à minima semi-automatique est indispensable. Un tel système, comme proposé en [De la Torre and Cohn, 2011] réalise automatiquement la détection des Action Units, mais sous la supervision d'un opérateur qui doit toujours marquer manuellement le point de départ de la séquence et le point d'apex. La solution

La subjectivité des labels est un second point. Plus la place laissée à la subjectivité du codeur est importante, plus le nombre de codeurs est amené à augmenter pour obtenir un taux d'accord acceptable.

Avec des systèmes limités aux 6 émotions de base tels que ceux qui existent aujourd'hui, les taux d'accord son élevés, mais ceci repose entièrement sur la théorie des émotions de base et leur lien codé « en dur »dans le cerveau humain.

On retombe sur la contradiction relevée précédemment, à savoir que pour réconcilier la technique avec la théorie psychologique, il faudrait pouvoir annoter une séquence sur une échelle continue dans le temps, y compris dans les phases où aucun signal n'est diffusé.

Par exemple, de multiples hochements de tête dans une situation réelle appréhendée par un humain, marquent l'approbation. Mais ils ne sont pas continus, or la séquence d'approbation elle, ne cesse pas entre 2 mouvements de tête.

Un algorithme de machine learning configuré pour inférer sur des labels discrets ne peut pas prendre en compte cette dimension temporelle.

Enfin, les biais dans les datasets eux-mêmes existent. Les équipes de recherchent développent et testent leurs solutions sur des bases de données publiques et ouvertes. Ils peuvent contenir des biais dès le départ.

Ils n'ont pas été conçus par exemple, pour être représentatifs ni de la diversité ethnique, ni sociale ou démographique. L'exploitation de biais (en machine learning, des hypothèses faites en amont qui permettent par exemple d'ignorer une partie des données) peut avoir un effet positif sur la performance, mais limiter la généralisation.

Le risque serait donc d'orienter la recherche vers une maximisation de la performance en recourant à des datasets très spécifiques, qui limitent la généralisation.

On peut citer comme exemple l'hypothèse d'avoir toujours, dans les datasets, des séquences segmentées, et non entrelacées, le fait d'avoir dans chaque séquence, au moins une expression neutre, ou l'éclairage contrôlé. « Fine-tuner »les données de test en les pré-segmentant de la sorte ne nous dit plus rien de ce qu'ils vaudraient en condition réelle.

Pour terminer, nous avons tenté, dans ce mémoire d'explorer la reconnaissance faciale pour donner au lecteur un angle plus large que la seule problématique du traitement de l'image et de la classification par un algorithme de machine learning. Nous avons voulu en éclairer la construction théorique, car il nous semble qu'une technologie qui a de bonnes chances d'être utilisée un jour à des fins de contrôle et de prédiction de nos comportements se doit d'être questionnée. Ce qui est d'autant plus vrai d'une technologie basée sur le cadre posé par une seule personnalité, fut-elle aussi réputée que Paul Ekman. Nous espérons que l'illustration de ce que sciences informatiques et sciences humaines ne sont pas hermétiques l'une à l'autre aura éveillé chez chacun autant d'intérêt à la lecture que pour nous à l'écriture.

Bibliographie

Ambady, N. and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences : A meta-analysis. *Psychological bulletin*, 111(2) :256.

Arnold, M. B. (1960). Emotion and personality.

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5) :1161–1179.

Barrett, L. F. (2006). Are emotions natural kinds ? *Perspectives on psychological science*, 1(1) :28–58.

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J. (2006). Fully automatic facial action recognition in spontaneous behavior. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 223–230. IEEE.

Bechara, A. (2004). The role of emotion in decision-making : evidence from neurological patients with orbitofrontal damage. *Brain Cogn.*, 55(1) :30–40.

Berkowitz, L. (1993). Towards a general theory of anger and emotional aggression : Implications of the cognitive-neoassociationistic perspective for the analysis of anger and other emotions.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., and Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE) : Reliability Data and Cross-National Differences. *J. Nonverbal Behav.*, 21(1) :3–21.

Bjørnsten, T. B. and Zacher Sørensen, M.-M. (2017). Uncertainties of facial emotion recognition technologies and the automation of emotional labour. *Digital Creativity*, 28(4) :297–307.

Brosch, T., Scherer, K. R., Grandjean, D., and Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Med. Wkly*, 143 :w13786.

Calvo, R. A. and D'Mello, S. (2010). Affect Detection : An Interdisciplinary Review of Models, Methods,and Their Applications. *IEEE Transactions on Affective Computing*, 1(1) :18–37.

Castellano, G., Caridakis, G., Camurri, A., Karpouzis, K., Volpe, G., and Kollias, S. (2010). Body gesture and facial expression analysis for automatic affect recognition. *Blueprint for affective computing : A sourcebook*, pages 245–255.

Ceyhan, A. (2006). Enjeux d'identification et de surveillance à l'heure de la biométrie. *Cultures & conflits*, (64) :33–47.

Cohn, J. F. (2006). Foundations of Human Computing : Facial Expression and Emotion. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, ICMI '06, pages 233–238, New York, NY, USA. ACM.

Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system.

Cohn, J. F., Zlochower, A. J., Lien, J., and Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, 36(1) :35–43.

Coppin, G. and Sander, D. (2010). *Théories et concepts contemporains en psychologie de l'émotion*, pages 25–56. Systèmes d'interaction émotionnelle. Hermès Science publications-Lavoisier, Paris. ID : unige :34368.

Damasio, A. R. (2003). *Spinoza avait raison*. Odile Jacob.

Damasio, A. R. (2006). *Erreur de Descartes (L')*. Odile Jacob.

Darwin, C. (1872). *The expression of the emotions in man and animals*.

De la Torre, F. and Cohn, J. F. (2011). Facial Expression Analysis. In Moeslund, T. B., Hilton, A., Krüger, V., and Sigal, L., editors, *Visual Analysis of Humans : Looking at People*, pages 377–409. Springer London, London.

Dickerson, S. S. and Kemeny, M. E. (2004). Acute stressors and cortisol responses : a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3) :355.

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion". *The Journal of General Psychology*, 25(2) :283–293.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press.

Ekman, P. (1982). Methods for measuring facial action. *Handbook of methods in nonverbal behavior research*, pages 45–90.

Ekman, P. (2007). *Emotions revealed : Recognizing faces and feelings to improve communication and emotional life*. Macmillan.

Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior : Categories, origins, usage, and coding. *semiotica*, 1(1) :49–98.

Ekman, P. and Friesen, W. V. (1975). Unmasking the face : A guide to recognizing emotions from facial cues.

Ekman, P. and Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.

Frank, M. and Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of personality and social psychology*, 80(1) :75–85.

Frank, M. G. and Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of personality and social psychology*, 86(3) :486.

Fridlund, A. J. (1994). *Human facial expression : An evolutionary view*. Academic Press.

Frijda, N. (1989). *Les émotions(pp.21-72)*. Neuchâtel-Paris : Delachaux & Niestlé.

Frijda, N. H. (1986). The emotions : Studies in emotion and social interaction. *Paris : Maison de Sciences de l'Homme*.

Frijda, N. H. (1988). The laws of emotion. *American psychologist*, 43(5) :349.

Gendron, M., Mesquita, B., and Barrett, L. F. (2013). Emotion Perception : Putting the Face in Context. In Reisberg, D., editor, *The Oxford Handbook of Cognitive Psychology*. Oxford University Press.

Goldberg, H., Preminger, S., and Malach, R. (2014). The emotion-action link ? Naturalistic emotional stimuli preferentially activate the human dorsal visual stream. *Neuroimage*, 84 :254–264.

Goldhaber, M. H. (1997). The attention economy and the net. *First Monday*, 2(4).

Grandjean, D. M. and Scherer, K. R. (2009). *Théorie de l'évaluation cognitive et dynamique des processus émotionnels*. Traité de Psychologie des émotions. Dunod, Paris. ID : unige :96517.

Griffiths, P. E. (2004). Towards a machiavellian theory of emotional appraisal. *Emotion, evolution and rationality*, pages 89–105.

Haidt, J. and Keltner, D. (1999). Culture and facial expression : Open-ended methods find more expressions and a gradient of recognition. *Cognition & Emotion*, 13(3) :225–266.

Hudlicka, E. (2015). *Computational Analytical Framework for Affective Modeling : Towards Guidelines for Designing Computational Models of Emotions*. IGI Global.

Hume, D. (1978). A treatise of human nature, 1739, 2d edition of 1888, edited by la selby bigge, revised by ph nidditch.

Ioannou, S. V., Raouzaiou, A. T., Tzouvaras, V. A., Mailis, T. P., Karpouzis, K. C., and Kollias, S. D. (2005). Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Netw.*, 18(4) :423–435.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science*, 2(3) :260–280.

Izard, C. E. (2008). Emotion Theory and Research : Highlights, Unanswered Questions, and Emerging Issues. *Annu. Rev. Psychol.*, 60(1) :1–25.

Izard, C. E. (2009). Emotion theory and research : Highlights, unanswered questions, and emerging issues. *Annual review of psychology*, 60 :1–25.

James, W. (1884). What is an emotion ? *Mind*, 9(34) :188–205.

Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53.

Kessous, E., Mellet, K., and Zouinar, M. (2010). L'économie de l'attention : entre protection des ressources cognitives et extraction de la valeur. *Sociologie du travail*, 52(3) :359–373.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., and Jain, A. K. (2012). Face Recognition Performance : Role of Demographic Information. *IEEE Trans. Inf. Forensics Secur.*, 7(6) :1789–1801.

Lazarus, R. S. (1966). Psychological stress and the coping process.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American psychologist*, 46(4) :352.

Lerner, J. S. and Keltner, D. (2000). Beyond valence : Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, 14(4) :473–493.

Lerner, J. S. and Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology*, 81(1) :146.

Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. *Annu. Rev. Psychol.*, 66 :799–823.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE.

Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J., and Budynek, J. (1998). The japanese female facial expression (jaffe) database. In *Proceedings of third international conference on automatic face and gesture recognition*, pages 14–16.

Marsella, S., Gratch, J., and Petta, P. (2010). Computational models of emotion. *A Blueprint for Affective Computing : A Sourcebook and Manual*.

Mehrabian, A. (1996). Pleasure-arousal-dominance : A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4) :261–292.

Mehrabian, A. (2008). Communication without words. *Communication theory*, pages 193–200.

Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal Theories of Emotion : State of the Art and Future Development. *Emot. Rev.*, 5(2) :119–124.

Mortillaro, M., Meuleman, B., and Scherer, K. R. (2012). Advocating a Componential Appraisal Model to Guide Emotion Recognition. *International Journal of Synthetic Emotions (IJSE)*, 3(1) :18–32.

Niedenthal, P. M., Krauth-Gruber, S., and Ric, F. (2009). *Comprendre les émotions : perspectives cognitives et psycho-sociales*. Wavre : Belgique.

Nugier, A. (2009). Histoire et grands courants de recherche sur les émotions. *Revue électronique de psychologie sociale*, 4(4) :8–14.

Pantic, M. and Bartlett, M. S. (2007). Machine analysis of facial expressions. In *Face recognition*. InTech.

Pantic, M. and Rothkrantz, L. (2004). Case-based reasoning for user-profiled recognition of emotions from face images. In *2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763)*, volume 1, pages 391–394 Vol.1.

Pantic, M. and Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expressions. *Image Vis. Comput.*, 18(11) :881–905.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *2005 IEEE International Conference on Multimedia and Expo*, pages 5 pp.–.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives ? *Pers. Soc. Psychol. Rev.*, 9(4) :278–311.

Picard, R. (1995). *Affective computing*. MIT Press, Cambridge, Mass.

Raghunathan, R. and Pham, M. T. (1999). All negative moods are not equal : Motivational influences of anxiety and sadness on decision making. *Organizational behavior and human decision processes*, 79(1) :56–77.

Rimé, B. (2009). *Le partage social des émotions*. Presses universitaires de France.

Russell, J. A. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion : dissecting the elephant. *Journal of personality and social psychology*, 76(5) :805.

Russell, J. A. and Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of personality and social psychology*, 38(2) :311.

Salzman, C. D. and Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual review of neuroscience*, 33 :173–202.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., and Parrott, D. J. (2001). A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression. *J. Nonverbal Behav.*, 25(3) :167–185.

Schachter, S. and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5) :379.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion : Theory, methods, research*, 92(120) :57.

Scherer, K. R. and Ekman, P. (2014). *Approaches to emotion*. Psychology Press.

Scherer, K. R. and Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, 22(5) :789–801.

Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., and Huang, T. S. (2007). Authentic facial expression analysis. *Image Vis. Comput.*, 25(12) :1856–1863.

Tian, Y.-L., Kanade, T., and Cohn, J. F. (2005). Facial Expression Analysis. In *Handbook of Face Recognition*, pages 247–275. Springer New York.

Tomkins, S. S. (1962). *Affect Imagery Consciousness : The Positive Affect*.

Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Cowie, R., and Douglas-Cowie, E. (2002). Emotion recognition and synthesis based on mpeg-4 faps. *MPEG-4 Facial Animation*, pages 141–167.

Valstar, M., Pantic, M., Ambadar, Z., and Cohn, J. (2006). Spontaneous vs. posed facial behavior : Automatic analysis of brow actions. In *8th International Conference on Multimodal Interfaces, ICMI 2006*. ACM.

Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., and Scherer, K. (2011). The first facial expression recognition and analysis challenge. In *Face and Gesture 2011*, pages 921–926.

Valstar, M. F. and Pantic, M. (2006). Biologically vs. Logic Inspired Encoding of Facial Actions and Emotions in Video. In *2006 IEEE International Conference on Multimedia and Expo*, pages 325–328.

Valstar, M. F. and Pantic, M. (2012). Fully automatic recognition of the temporal phases of facial actions. *IEEE Trans. Syst. Man Cybern. B Cybern.*, 42(1) :28–43.

Wundt, N. (1897). *"Outlines of psychology"*. Oxford, England : Engelman.

Zajonc, R. B. (1984). On the primacy of affect.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods : audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(1) :39–58.